Department of Psychology and Logopedics,
Faculty of Medicine, Doctoral Programme Brain and Mind,
University of Helsinki

CONNECTIONS BETWEEN ARTICULATIONS AND GRASPING

Mikko Tiainen

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Medicine of the University of Helsinki, in Auditorium I, Metsätalo, Unioninkatu 40, on the 6th of August 2020 at 12 o’clock.

The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

ISBN 978-951-51-6283-0 (pbk.)
ISBN 978-951-51-6284-7 (PDF)

Unigrafia
Helsinki 2020

Supervisors	Docent Kaisa Tiippana Department of Psychology and Logopedics Faculty of Medicine University of Helsinki Finland
	Docent Lari Vainio Helsinki Collegium for Advanced Studies University of Helsinki Finland
	Professor Martti Vainio Department of Digital Humanities Faculty of Arts University of Helsinki Finland
Reviewers	Directeur de Recherche Jean-Luc Schwartz Le laboratoire Grenoble Images Parole Signal Automatique Le Centre National de la Recherche Scientifique France
	Professor Elina Mainela-Arnold Department of Psychology and Speech-Language Pathology University of Turku Finland
Opponent	Assistant Professor Riikka Möttönen School of Psychology University of Nottingham United Kingdom

ABSTRACT

The idea that hand gestures and speech are connected is quite old. Some of these theories even suggest that language is primarily based on a manual communication system. In this thesis, I present four studies in which we studied the connections between articulatory gestures and manual grasps. The work is based on an earlier finding showing systematic connections between specific articulatory gestures and grasp types. For example, uttering a syllable such as [kɑ] can facilitate power grip responses, whereas uttering a syllable such as [ti] can facilitate precision grip responses. I will refer to this phenomenon as the articulation-grip congruency effect. Similarly, to the original work, we used special power and precision grip devices that the participants held in their hand to perform responses. In Study I, we measured response times and accuracy of grip responses and vocalisations to investigate whether the effect can be also observed in vocal responses, and to which extent the effect operates in the action selection processes. In Study II, grip response times were measured to investigate whether the effect persists when the syllables are only heard or read silently. Study III investigated the influence of grasp planning and/or execution on categorizing perceived syllables. In Study IV, we measured electrical activity in the brain during listening of syllables that were either congruent or incongruent with the precision or power grip, and we investigated how performing different grips affected the auditory processing of the heard syllables.

The results of Study I showed that besides manual facilitation, the effect is observed also in vocal responses, both when a simultaneous grip is executed and when it is only prepared, meaning that overt execution is not needed for the effect. This suggests that the effect operates in action planning. In addition, the effect was also observed when the participants knew beforehand which response they should execute, suggesting that the effect is not based on the action selection processes. Study II showed that the effect was also observed when the syllables were heard or read silently, supporting the view that articulatory simulation of a perceived syllable can activate the motor program of the grasp which is congruent with the syllable. Study III revealed that grip preparation can influence categorization of perceived syllables. The participants were biased to categorize noise-masked syllables as being [ke] rather than [te] when they were prepared to execute the power grip, and vice versa when they were prepared to execute the precision grip. Finally, Study IV showed that grip performance also modulates early auditory processing of heard syllables.

These results support the view that articulatory and hand motor representations form a partly shared network, where activity from one domain

can induce activity in the other. This is in line with earlier studies that have shown more general linkage between mouth and manual processes and expands this notion of hand-mouth interaction by showing that these connections can also operate between very specific hand and articulatory gestures.

TIIVISTELMÄ

Ajatus käden eleiden ja puheen välisistä yhteyksistä on melko vanha. Jotkut teoriat jopa ehdottavat, että kieli pohjautuu pääosin käsillä tapahtuvaan kommunikointijärjestelmään. Tässä väitöskirjassa esittelen neljä osatyötä, joissa tutkimme artikulatoristen eleiden ja tarttumisotteiden välisiä yhteyksiä. Työ perustuu aiempaan löydökseen, joka paljasti systemaattisia yhteyksiä tiettyjen artikulatoristen eleiden ja tarttumisotteiden välillä. Esimerkiksi [kɑ] tavun lausuminen nopeuttaa voimaotteen tekemistä, kun taas esimerkiksi [ti] tavun lausuminen nopeuttaa pinsettiotteen tekemistä. Väitöskirjan osatyöt hyödynsivät tätä perusefektiä muokkaamalla koeasetelmaa kuhunkin tutkimuskysymykseen sopivaksi.

Osatyön I tulokset osoittivat, että yhteensopivuusefekti on havaittavissa myös lausutuissa vastauksissa. Efekti havaittiin myös, kun otteen suorittamiseen oli vain valmistauduttu. Tämä viittaa siihen, että efekti toimii toimintojen suunnittelun tasolla. Lisäksi efekti havaittiin silloinkin, kun osallistujat tiesivät etukäteen, mikä vastaus heidän tulisi suorittaa, mikä viittaa siihen, ettei efekti perustu toimintojen valintaan liittyviin prosesseihin. Osatyössä II efekti havaittiin, vaikka tavut vain kuultiin tai luettiin äänettömästä. Tämä tukee näkemystä, että havaittujen tavujen artikulatorinen simulointi voi aktivoida tavun kanssa yhteensopivan otteen motorista ohjelmaa. Osatyö III osoitti, että käden otteet voivat vaikuttaa havaittujen tavujen luokitteluun. Osallistujat olivat biasoituneet luokittelemaan esitettyjen kohinaisten tavujen olevan ennemmin [ke] kuin [te], kun he olivat valmistautuneet suorittamaan voimaotteen ja päinvastoin, kun he olivat valmistautuneet pinsettiotteen suorittamiseen. Viimeisimpänä osatyö IV osoitti, että otteiden suorittaminen vaikuttaa myös havaittujen tavujen varhaiseen auditoriseen prosessointiin.

Nämä tulokset tukevat näkemystä, että artikulatoriset ja käden motoriset edustukset muodostavat osittain jaetun verkoston, jossa aktiivisuus yhdellä osa-alueella voi aiheuttaa aktiivisuutta myös toisella. Tämä on linjassa aiheen aiempien tutkimusten kanssa, jotka ovat osoittaneet yleisempiä yhteyksiä käden ja suun toimintojen välillä. Nämä tulokset laajentavat käden ja suun välisen yhteyden ajatusta osoittamalla, että yhteydet voivat toimia myös hyvin tarkasti rajattujen artikulatoristen ja käden eleiden välillä.

ACKNOWLEDGEMENTS

"You're a wizard Harry!" I may not be Harry Potter, but I do feel a bit like a wizard now that I am finally finishing this thesis. This has been a long journey. There are many people that I want to give my thanks to. First of all are my three supervisors: Kaisa Tiippana, Lari Vainio and Martti Vainio. Thanks for taking me into the Mind-project. I am sorry that this dragged on for so long and caused you all stress as well. Kaisa, you supervised my first bachelor's thesis and my master's thesis, and now my PhD. You enabled me to pursue this career of research and taught me a lot about how to do research and especially how to write scientific texts and how to keep them concise. I hope I have improved since the early days. Lari, we have set up many studies together and collected a lot of data. It has always been the most enjoyable part of doing research for me. You are truly an expert of this field and I could never hope to call myself an expert knowing how well-versed you are in the literature. Martti, I have always admired and enjoyed your enthusiastic attitude and openness to trying new things. I also feel that a big part of why I am a software developer today is because of your encouragement to learn to code and the go-ahead for building new experiment setups with Arduinos.

I want to thank our lunch gang at Siltavuorenpenger: Viljami Salmela, Markku Kilpeläinen, Sointu Leikas, Jenni Heikkilä and all the others who joined on several occasions. This was always a highlight of my day and I hope all of you had fun as well. Thanks to all the people who worked with me along the way on various experiments: Tarja Peromaa, Naeem Komeilipoor, Petri Paavilainen, Juraj Šimko, Jiří Lukavský, Fatima Felisberti, Aleksi Rantala, Tommi Makkonen and others who were involved.

Thank you to my preliminary examiners Jean-Luc Schwartz and Elina Mainela-Arnold for the kind words and valid criticism that you gave. Thank you to Riikka Möttönen for agreeing to be my opponent and Kimmo Alho for agreeing to be the custos. Thank you also to the Academy of Finland and Alfred Kordelin Foundation for funding my PhD process.

Like I said, this has been a long journey. I started as a PhD student in May 2013 and it is now June 18th 2020 when I'm writing this. Since then a lot of things have happened. There have been ups and downs. Most notably when it was time to write this thesis. In 2017 I was working on it along with my computer science studies. Writing was not progressing too well, and I felt a constant writer's block. I had set a hard deadline for myself for the thesis to be done at the end of the year since I was starting at a new job. I kept thinking that I just need a couple weeks to write this; I just need to push it a bit more. This has usually been my strategy; when there is a lot of work, I just work a lot.

Finally, on 23rd of December I realized that it was not going to happen. I was completely burned out. In hindsight I should have noticed the symptoms already during the spring. But that is just one of the symptoms that you become blind to them. Also, in hindsight it was a bad idea to progress the computer science studies so heavily during writing. But then again, because of that I got the new job, which led to the next job and to the path that I am on right now, so I cannot say it was all bad.

A big thank you then goes to my then-new boss, Matti Luukkainen, who was very understanding when I told him about the burnout and he allowed me to take it easy for the beginning of the year. This was immensely helpful. So, slowly I started to recover. I did not open the thesis file for almost a year. But after a year, I was able to slowly get back to the text. It still caused quite a bit of anxiety, but slowly it started moving forward. Thanks also to Arto Hellas for the discussions we had about it during car rides. They helped me to understand that I should write the thesis with the general public in mind which helped me in finding the tone for the text. Thanks to Juha Tauriainen for the peer support when you were writing your own bachelor's thesis. Thank you to my dear friend Andrei Luukka for also always being there to listen to my head aches.

Big thanks to all other friends and family as well, especially everyone who kept asking "When is the PhD done?" That never got old during the seven years. But seriously, thank you for showing interest. Last but not least, huge thanks to my lovely wife, Leena Tiainen who has stuck with me through it all and always stayed supportive and encouraging.

I decided a long time ago that the acknowledgements section will be the last thing I write. I do love doing research, it has been great, but now it is time for something else for a change. Thanks.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications:

I Tiainen, M., Tiippana, K., Vainio, M., Komeilipoor, N., & Vainio, L. (2017). Interaction in planning vocalizations and grasping. *The Quarterly Journal of Experimental Psychology, 70(8),* 1590–1602.

II Vainio, L., Tiainen, M., Tiippana, K., & Vainio, M. (2014). Shared processing of planning articulatory gestures and grasping. *Experimental Brain Research, 232(7)*, 2359–2368.

III Tiainen, M., Tiippana, K., Vainio, M., Peromaa, T., Komeilipoor, N., & Vainio, L. (2016). Selective influences of precision and power grips on speech categorization. *PLOS One. 11(3)*, e0151688.

IV Tiainen, M., Tiippana, K., Paavilainen, P., Vainio, M., & Vainio, L. (2017). Mismatch negativity (MMN) to speech sounds is modulated systematically by manual grip execution. *Neuroscience Letters, 651,* 237–241.

ABBREVIATIONS

AGC	articulation-grip congruency
EEG	electroencephalography
EMG	electromyography
fMRI	functional magnetic resonance imaging
MEG	magnetoencephalography
MEP	motor evoked potentials
MMN	mismatch negativity
TMS	transcranial magnetic stimulation

1 INTRODUCTION

This thesis is about grasping. Grasping is a very fundamental human action that we perform countless times every day. This thesis is also about articulation, or articulatory gestures, another fundamental human action. How are these actions connected, and why? That question is at the core of this thesis. This is by no means a novel question. Darwin (1872) already speculated about the connections between hand and mouth actions by how children tend to twist their tongue when they are learning to write or how people tend to move their jaw rhythmically when cutting with scissors. The latter example is interesting in that it is an example of a non-language related connection between mouth and hand actions.

These kinds of connections are what I will be focusing on in this thesis, more precisely how the non-communicative grasping gestures are connected to articulatory gestures of the mouth. I will, however, start by going over communicative gestures more generally and discussing theories on how human communication and language has evolved, at the same time narrowing the scope to the specific gestures of grasping and articulating. Finally, I'll explain our previous findings about how grasping and articulations actually are connected, before moving onto the specific studies of this thesis.

1.1 WHAT ARE GESTURES?

The most apparent manifestation of the link between speech and hand movements are co-speech gestures. These gestures are an integral part of our everyday life. They are used in many different ways to communicate our intentions to others. These gestures are also the most researched expression of the link between speech and hand movements (McNeill, 1992). They have been mostly explored by analyzing overt hand movements during communication between individuals. I will now briefly go through different types of co-speech gestures that exist. Although these kinds of co-speech hand gestures are not the focus of this thesis, they are the most explicit way people encounter the hand and mouth movement connections and also offer some insights about the nature of the connections between hand and mouth movements. They may also be based on the same mechanisms as the connections between mouth actions and the non-communicative grip gestures (see Vainio, 2019 for a review) which are the main focus of this thesis.

The co-speech gestures can be categorized in different ways. McNeill (1992), for example, divides them into gesticulations, speech-framed gestures, emblems, pantomimes and signs. Gesticulations are usually made with arms and hands in synchrony with the accompanying speech. They are motions that embody a meaning relatable to the speech. Describing the size of a fish you caught and at the same time showing it with your hands would be one example of a gesticulation. Speech-framed gestures are gestures that fill in the actual sentence. Thus, unlike gesticulations, they are not synchronous with speech, but performed sequentially with speech. Making a suggestion like "Would you like to go get a couple..." and follow it with a drinking motion with the hand, for example, could be considered a speech-framed gesture where the hand motion completes the suggestion to go for a drink. Emblems are culturally specific, conventionalized signs that are meaningful even without speech. A well-known western example of an emblem would be the thumbs-up sign signalling approval. Pantomimes are gestures or sequences of gestures that tell a story, without overt speech. Signs are words in sign languages that have their own linguistic structures. McNeill (1992) describes this division of gestures as a continuum, where the importance/involvement of speech decreases when moving from gesticulations to signs.

Gesticulations are further divided into iconic, metaphoric, deictic and beat categories. Iconic gestures are illustrations of concrete actions or events. In contrast, metaphoric gestures depict abstract constructs as if they were something concrete. Deictic gestures are usually done by pointing at something (concrete or abstract) with the index finger, but other body parts can also be used. Beats are gestures where the hands are moving in the rhythm of the speech.

This is not the only way to categorize gestures, De Ruiter (2000), for example, divides gestures slightly differently into iconic gestures, deictic gestures, beat gestures, pantomimes and emblems. Regardless of how we categorize gestures, they are clearly a major part of our communication and in many ways connect with speech. It has been suggested that this connection between gestures and speech could be based on common evolutionary history between the two (e.g. Rizzolatti & Arbib, 1998), and could be related to the actual evolution of spoken language. Theories that follow this train of thought are commonly referred to as "gestural theories of language evolution" (e.g., Hewes 1973; Arbib, 2005; Gentilucci & Corballis, 2006; Vainio, 2019).

These co-speech gestures such as emblems and pantomimes are purely performed for communicative purposes, and they demonstrate in very explicit manner how manual movements are an integral part of speech. However, many gestural theories of language evolution assume that initially the primary basis for speech evolution was built on manual representation that involve hand movements for other purposes than communication, such as grasping and manual manipulation (e.g., Hewes 1973; Arbib, 2005). In addition, even though these co-speech gestures show a visible link between speech and manual movements, they are mostly linked to semantic (e.g., iconic gestures, pantomimes and emblems) and prosodic (e.g., beat gestures) aspects of speech. Regarding speech evolution, however, the core element that needed to be developed was articulatory gestures. Gestural theories of language evolution largely emphasize that utilization of particularly those manual representations in speech domain that were not directly related to communication – such as grasping – facilitated development of articulatory gestures. In other words, according to these theories, it is possible that for example the emergence of some consonants might have been triggered by adapting manual grasp representations for shaping articulatory gestures.

1.2 THE EVOLUTION OF LANGUAGE

There are many different theories about the origins of language, one of the most famous ones is probably Noam Chomsky's single step theory (e.g., Chomsky, 2005). This theory suggests that language emerged rather quickly, as a kind of innate, universal grammar, which together with environmental exposure forms into specific languages. This is in sharp contrast to the gestural theories of language evolution. Gestural theories usually suggest that language evolved gradually from gestural communication to spoken languages (e.g., Rizzolatti & Arbib, 1998; Corballis & Gentilucci, 2006). They also reject the idea of universal or innate grammar, and rather suggest that language is built on the human capability to understand the intentions of other people's actions. Some of the gestural theories emphasize that communication occurred initially via gestures of hand and body before starting to use mouth as a primary source of communication (e.g., Arbib, 2005), while other gestural theories do not take a strong stance on whether communication occurred initially vial body gestures, and rather highlight that speech might have facilitated shaping of articulatory gestures during early stages of speech development (e.g., Vainio, 2019). Nevertheless, all different views of gestural theories underline the tight connection between manual actions and speech evolution.

Hauser and Fitch (2003) suggest that speech evolved by taking advantage of structures or functions not originally developed for speech or language. For example the lowered larynx of humans might have initially evolved to convey size information about the individual (lower larynx enables lower-pitched vocalizations which is associated with a larger individual). This, with other improved vocal capabilities would then later have been adopted to use in speech and allowed for a wide range of formant patterns.

Another quite well-known theory following a somewhat similar vein as Hauser and Fitch (2003) is MacNeilage's (1998) frame/content theory of language evolution. The main idea of this theory is that language evolved from ingestive mouth actions. The suggestion is that the system that controls the rhythm of our speech is based on the processes that control the rhythm of ingestive mouth actions (e.g. chewing, sucking) . The open-close alternation naturally constitutes the vowel-consonant structure of syllables, respectively. MacNeilage (1998) proposes that the ingestive actions could have first taken the shape of communicative gestures as lip-smacks and teeth chatters etc., similar to what can be readily observed in nonhuman primates. These communicative mouth gestures of monkeys have actually been proposed by Van Hoof (1967) to be ritualizations of ingestive actions that are linked to social bonding. For example, when a monkey is grooming another one, it eats the fleas it finds and this produces lip-smacking. From this, lip-smacking has become a generalized communicative gesture that can be used even if the monkeys are not grooming one another and they can both still understand the meaning of it. The "frame" of MacNeilage's (1998) theory refers to the simple opening and closing of the mouth, the principle component of speech, which develops first. The "content" refers to the ability of humans to use our articulators to form complex vowel and consonants sounds. This control develops later during language developments of an infant, and sometimes is left incomplete, as is the case in speech sound disorders (MacNeilage, 1998).

I will focus mainly on gestural theories of language evolution, as they are closely related to the main topic of this thesis. It is, however, important to note that these are by far not the only possible way language has evolved, of which the above one's are just a couple examples. Now, getting back to the idea of gestural origin of language, it is actually a rather old idea. As mentioned earlier, the connections between hand and mouth actions were already noted by Darwin (1872). More direct suggestions about the evolution of language from gestures came e.g. from Sweet (1888) and Wallace (1881), who suggested that articulations became linked with hand gestures by directly combining articulations with hand gestures or by the articulators roughly imitating hand (or other bodily) gestures.

For a more comprehensive review of the theories prior to the 20th century, see Woll, (2014) and Wacewicz, Żywiczyński and Orzechowski (2016), but let us here take a bit closer look at the theories by Paget (1930) and Hewes (1973). Paget (1930) suggested that speech in general was perceived through gestures that produce the sounds rather than the absolute sounds themselves. This would explain how we are able to perceive two words as being the same when uttered by, for example, a male speaker and a female speaker, even though the actual sound spectrum can be quite different. This idea is very similar to the motor theory of speech perception (Liberman., Cooper, Shankweiler & Studdert-Kennedy, 1967) which I will get back to in more detail in Study IV. Paget (1930) further speculated that this speech based on articulation gestures was originally developed from using whole-body gestures. He contrasts this with other primates, where "lower-level" monkeys are louder, using more calls related to emotional state, whereas "higher-level" apes are more silent and communicate more through gestures. Similarly, Paget (1930) suggests that our ancestors first communicated our thoughts to others through gestures, such as those described by McNeill (1992) above. Why humans transitioned to a speech-based communication system, was not because of a need for more expressiveness but more because of pure convenience when humans had their hands full with increasing crafting and tool work (Paget, 1930). Now, as Darwin (1872) had already noted that humans tend to mimic hand gestures with the mouth, since the hands were occupied with crafting, the mouth that had mimicked the communicative hand gestures as a secondary source, gradually became the primary gesturing source (Paget, 1930). Add to this the discovery that adding airflow during an articulatory gesture produces sounds from which the actual gesture can be deciphered, and you have first stages of spoken language (Paget, 1930).

Woll (2014) describes the gestural theory proposed by Hewes (1973) as a bridge between the old and contemporary gestural theories of language evolution. Hewes (1973), like Paget (1930), emphasized that the emotion-based vocalisations of monkeys are not as likely to be the basis for modern language as is the more cognitively demanding gestural communication. He described imitation of other animals and their gestures as an important step on the road to acquiring language. Imitation is also how our ancestors (and humans today) learned tool use, and imitating a tool's usage might have become a gesture to mean that tool (Hewes, 1973). We will come back to imitation and its importance a bit later on. Nevertheless, Hewes (1973) admits that there are still many question marks about how communication moved from hand-based gestures to mouth-based. He does bring up some points about this such as the existing connection between hands and mouth in the form of eating (Hewes, 1973), which could possibly help moving from one modality to another. He also mentions the tendency of humans to explore things by putting them in their mouth (Hewes, 1973). He points in the direction of the earlier mouth-gesture hypotheses, such as Paget's (1930) work, and how these theorize that humans use the mouth and lips to imitate hand movements. Although the reason 'how' is quite unclear, Hewes (1973) does list a variety of reasons 'why'. Like Paget (1930) mentioned, it freed the hands for tool use, but also speech enabled communication in the dark and over longer distances, and in general removed the need for visual contact (Hewes, 1973).

Gestural theories of language evolution have seen a resurgence during the past twenty years (e.g. Rizzolatti & Arbib, 1998; Arbib, 2005; Corballis & Gentilucci, 2006). One of the reasons for this renaissance is the discovery of mirror neurons (Di Pellegrino, Fadiga, Fogassi, Gallese & Rizzolatti, 1992). These neurons were discovered in the macaque monkey premotor cortex area F5, which is considered to be the homologue of Broca's area in humans (Rizzolatti & Craighero, 2004). Broca's area is well-known to be crucial for speech, and damage to this area can result in expressive aphasia (also known as Broca's aphasia), which includes trouble producing speech. Di Pellegrino et al. (1992) originally observed that there were neurons in this area F5 that discharge both when a monkey performs a goal-directed hand action, such as grasping a raisin, and when the monkey simply observes someone else perform that same action. Since then, mirror neurons and the mirror system have become a significant research area, and evidence of similar neurons in humans have been accumulating, (e.g. Mukamel, Ekstrom, Kaplan, Iacoboni & Fried, 2010; see Rizzolatti & Craighero, 2004 for a review of different studies of the mirror system). Mirror neurons are thought to be a key factor in human imitation capabilities (e.g. Rizzolatti & Craighero, 2004). Since these neurons are active both when perceiving an action and performing it, they can bridge the gap of performing perceived actions by mapping them directly to our own motor repertoire. Basically, when you see someone swinging a baseball bat, the mirror neurons can help you understand how you would have to move your own body in order to perform that same maneuver. Today, the mirror system is considered by many to be a way for humans to understand the intentions of others, and a system by which we convey our intentions to others. Since the role of imitation was so important in the earlier gestural theories of Hewes (1973) and Paget (1930), it is now probably obvious why gestural theories of language evolution have seen a resurgence after the discovery of mirror neurons, as they provide a neurological basis for imitation.

One important additional note about the mirror neurons is that the original mirror neuron studies considered only hand actions. However, Ferrari, Gallese, Rizzolatti and Fogassi (2003) recorded mirror neurons in macaque monkeys that respond to mouth actions. The majority of these neurons respond to performing and observing ingestive mouth actions (e.g. sucking), but some respond effectively to communicative mouth gestures, such as lip-smacking. The fact that mirror neurons are not limited to just hand actions but are also involved in decoding speech cues is important when we are talking about speech.

Now, let us turn towards the newer gestural theories, namely those by Rizzolatti and Arbib (1998), and Corballis and Gentilucci (2006). The idea of Rizzolatti and Arbib (1998) is that manual gestures were used to complement oro-facial gestures, for example by pointing at something with the finger and then performing a facial gesture (e.g., a lip protrusion) related to the pointed object. This increased the power of this kind of communication and at the same time increased the value of simultaneous vocalisations. This view assumes that pairing analogous vocalisations (e.g. mouth wide open vocalisations with spreading arms open or narrow mouth vocalisations with pinching something small with the fingers) simultaneously with the manual gestures could transfer the intent of the gestures to the vocalisations, reducing the importance of the manual gesture and leading to a primitive vocal grammar. Arbib (2005) further developed this theory and described different criteria of language readiness that humans needed to create a “true” language. However, after these criteria for language readiness are met, the theory does not necessitate that a language will automatically emerge, nor that it will be a spoken one. Further cultural evolution needs to happen for language development, but Arbib (2005) claims that no more neural structural changes are necessary. I will now discuss a bit more in depth this theory and especially what the criteria for language readiness are.

At the heart of this extended “Mirror system hypothesis” is, once again, the concept of imitation. Arbib (2005) claims monkeys, apes and humans can be distinguished by the level of imitation they can perform. According to this view, monkeys can only copy movements. For example, when marmosets see a researcher open a canister using their mouth to obtain a mealworm, the marmosets are more likely to subsequently use their mouth to open the canister lid when given the opportunity (Voelkl & Huber, 2000). Although Voelkl and Huber (2000) considered this as an imitation, Arbib (2005, p. 114) does not. Instead he labels it as “stimulus enhancement, apparent imitation resulting from directing attention to a particular object or part of the body or environment”. He contrasts this with true imitation: “which involves copying a novel, otherwise improbable action or some act that is outside the imitator’s prior repertoire” (Arbib, 2005, p. 114). This form of simple imitation is something that apes are capable of. Humans then are capable of what Arbib (2005) calls complex imitation, which involves being able to acquire longer, more complex and more abstract sequences of actions in a single trial, something that apes are not able to do. This capability for complex imitation is what Arbib (2005) considers to give rise to the ability to communicate via pantomimes, which includes both the ability to abstract actions, such as signaling a bird flying by flapping the hands, but also the ability to understand the meaning of these abstractions. This will in turn lead to the emergence of protosign and then to protospeech, which would have developed concurrently, leading eventually to actual speech (Arbib, 2005).

Before moving on, it would be a good moment to remind that the title of the thesis is “Connections between articulations and grasping”. If manual gestures in general are connected to language, why is this thesis focused on grasping actions? There is evidence which indicates that grasp actions could have a special connection with language. The theory of Corballis and Gentilucci (2006) agrees a lot with that of Rizzolatti and Arbib (1998; Arbib, 2005) but they draw special attention to grasping gestures. This is because it can be assumed that our ancestors had pre-existing connections between manual grasping and mouth actions in the form of eating behaviour, even before speech had evolved. This is closely in line with the theory of MacNeilage (1998), but Corballis and Gentilucci (2006) argue that it is only half of the story as the theory of MacNeilage (1998) fails to take into consideration the importance of hand movements in primate eating behaviours.

The tightness of this connection between hand and mouth actions is reaffirmed by the finding of neurons in macaque monkeys that are active both when the monkey grasps something with the hand and when it grasps something with the mouth (Rizzolatti et al., 1988). Corballis and Gentilucci (2006) build their account of language evolution on the idea that language evolved by utilising this already existing connection between hand and mouth actions. This view assumes that the switch from manual gesture communication to vocal-dominant communication would have been a gradual one. In the same vein, Corballis and Gentilucci (2006) describe how even now, in sign languages, mouth gestures can be used to disambiguate manual gestures. They actually consider manual and vocal communication a continuum, and support this view by the distinction that speech itself can be considered fundamentally gestural, as mentioned already above by Paget (1930) and in the motor theory of speech perception (Liberman et al., 1967). Taken these notions into consideration, especially how grasping is so important for primate eating behaviour that there are even neurons that react to grasping irrespective of whether it is done with the mouth or with the hand (Rizzolatti et al., 1988), grasping actions could present the strongest connections with speech.

1.3 DEVELOPMENTAL CONSIDERATIONS OF HAND-MOUTH CONNECTIONS

I should point out that the connections between language and hand actions does not require, nor can it prove, that language evolved from hand gestures. It is possible that language evolved in some other way, but the connections between language and gestures make it plausible that the evolution of language could be related to gestures. In their review, Willems and Hagoort (2007) approach the connection between language and gestures, and actions in general, from an embodied point of view, which, unlike the traditional Cartesian view, does not see mind and body as two separate "things". Rather, embodied cognition sees cognition (mind) arising from actions (body). A good example of this kind of embodied view of cognition is the motor theory of speech perception (Liberman et al., 1967; Liberman & Mattingly, 1985), which suggests that speech perception is based on mapping heard speech to one's own articulatory gestures. That is, speech perception is shaped by mapping the heard speech sounds to one's own articulatory motor actions (i.e., how one would make those sounds themselves).

In addition to evolutionary aspects of language, researchers have also proposed a tight link between speech development and manual action. For example, Greenfield (1991) suggests that in infant brain there is an undifferentiated circuit in Broca's area that is responsible for both hierarchical organization of language and manual object combination (e.g., tool use). During development this circuit becomes more differentiated and so the two functionalities also become differentiated from each other. Indeed, evidence suggests many ways development of hand and mouth actions in infants are coupled and can influence each other. For instance, nine-week old infants already display hand-mouth connections by being more likely to curl their fingers during vocalisation and to point with the index finger before or after vocalisation, whilst interacting with their mother (Fogel & Hannan, 1985). Another early example of hand-mouth connections is the Babkin-reflex of infants (Babkin, 1958). It can be elicited by pressing your thumbs against the palms of the infant while the infant is lying down. This results in the infant opening its mouth and also flexing its forearms and head, and closing its eyes (Futagi, Yanagihara, Mogami, Ikeda, & Suzuki, 2013). This reflex diminishes usually by the time the infant is four or five months old (Futagi et al., 2013), but a similar movement of hand-to-mouth can still be elicited in adults by electrically stimulating the precentral gyrus (Desmurget et al., 2014).

As infants learn to grasp objects, they quickly start moving the grasped objects to their mouth, an action commonly known as mouthing, and the frequency of this tendency quickly increases, peaking at around 7 months of age (Rochat, 1989). Mouthing has been proposed to play a part in speech development (Fagan & Iverson, 2007). When an infant puts an object in its mouth, the object closes the vocal tract and presses the tongue in different ways. If the infant then tries to produce sounds, it will result in different consonant sounds depending on how the tongue is pressed by the object in mouth (Fagan & Iverson, 2007). This age period is in general an important time in consonant development (Iverson, 2010), and also coincides with the emergence of vocal and manual babbling. In turn, manual babbling itself has been suggested to facilitate speech development (Iverson & Thelen, 1999).

There is also evidence which suggests that manual abilities can predict later speech development. For example, an infant's preference for using the right hand for grasping and object manipulation at 6-14 months, predicts relatively advanced language skill development at 24 months (Nelson, Campbell, & Michel, 2014). Similarly, a child's vocabulary on school entry at 54 months can be predicted by their manual gesturing at 14 months, where more gesturing predicts a larger vocabulary (Rowe & Goldin-Meadow, 2009). Poorer manual fine motor skills in childhood are in general associated with specific language impairments (Hill, 2001).

All these above mentioned evidence suggests that there is tight innate connections between mouth actions and manual actions of grasping and manipulation, and that these connections might have an important role in the development of speech. Above, I summarized some of the developmental evidence about the hand-mouth connections mainly from studies on infants. In the next section, I will introduce some of the evidence from studies on adults, both from brain imaging and behavioural studies. I will start with studies with primates and more general studies showing connections between grasping and non articulatory mouth movement and in the next section move on to those that deal with grasping and articulation, and thus are most relevant for this thesis.

1.4 EVIDENCE OF THE CONNECTION BETWEEN LANGUAGE AND GESTURES

As already mentioned, Rizzolatti et al., (1988) reported neurons that activate both when a monkey grasps an object with the hand or with the mouth. I will refer to these types of neurons hereafter as dual-grasp neurons. An interesting aspect of the results of Rizzolatti et al., (1988) is that some of these neurons show specificity for the type of hand grasping action performed. Some neurons are only active when a precision grip (thumb against index finger) is performed and some when a finger prehension grip (thumb against other fingers) is performed. They did not find neurons that were only active for whole hand grasping (or power grip; wrapping the fingers around the object). I will come back to different grip types a bit later. Gallese, Fadiga, Fogassi and Rizzolatti (1996) also reported mirror-like neurons that respond both when a monkey observes grasping actions performed with a hand or with the mouth. In line with this evidence, Waters and Fouts (2002) observed mouth movements (such as protrusion of the lips or tongue) of captive chimpanzees while performing manual manipulations. Of the times that the chimpanzees performed fine manual manipulations (e.g., precision grip), there were more times when they exhibited sympathetic mouth movements than those were they did not exhibit mouth movements. Contrarily, when the chimpanzees performed gross motor actions (e.g. power grip, or without grasping the object at all), there were more times when they did not exhibit simultaneous mouth movements than those when they did. Based on their results and those of Rizzolatti et al. (1988), Waters and Fouts (2002) deem it possible that when performing fine manual actions, the activity of those neurons spills over to mouth movements.

This all sounds quite a lot like the observations of Darwin (1872), mentioned earlier, about the human tendency to mimic scissor movement with the mouth. This leads us to the question, how have these connections been studied in humans? I will start with brain imaging/stimulation studies and move to behavioural evidence. Studies on humans have been conducted for example by using transcranial magnetic stimulation (TMS). In TMS, an electric coil is used to generate a magnetic field that can pass through the skull without damaging it and stimulate the neurons beneath the scalp. Stimulation can also be done directly, with electrodes placed right on the cortical surface. One example of such a study was already mentioned, the one done by Desmurget et al. (2014), where they found the effect that participants bring their hand to the mouth when stimulating the precentral gyrus, a result similar to the one Rizzolatti et al. (1988) found on monkeys.

Direct electronic stimulation is only possible on surgical patients due to its extremely invasive nature and can only be used when patients need the electrons placed to gather information for their upcoming surgery. Because of this, TMS is by far the more common stimulation method in human studies. When investigating the human motor processes, TMS is usually accompanied by electromyographic (EMG) measurements. EMG measures muscle activity, so a basic TMS study usually applies TMS to a motor brain area (e.g. hand) while simultaneously measuring EMG from the affected muscles (e.g. hand muscles). The activity induced by TMS in the hand muscles can then be observed in the EMG signal as motor evoked potentials (MEP). For example, Tokimura, Asakura, Tokimura, Oliviero and Rothwell (1996) used TMS to stimulate the hand motor areas and found that if participants are simultaneously reading aloud or speaking spontaneously, the EMG responses of hand muscles increase compared to a baseline condition. Meister et al., (2003) found a similar effect of increased hand muscle activity during reading aloud and also noted that no such increase was observed in foot muscles. These studies would thus indicate that speech can increase activity of hand muscles, and that this is specific to hands, and does not generalise to other effectors such as foot.

Simply listening to speech during TMS also increases the activity of hand muscles (Flöel, Ellger, Breitenstein & Knecht, 2003), although in a TMS study by our group, we only found activation of hand muscles during articulation and not while listening to speech (Komeilipoor, Tiainen, Tiippana, Vainio & Vainio, 2016). However, the speech stimuli we used were meaningless syllables, whereas Flöel et al. (2003) used full stories or short sentences, which could make a difference. Indeed, Flöel et al. (2003) found a trend that while listening to stories, activation of hand muscles tended to be greater than when listening to short sentences. It could be speculated that the naturalness of the stimuli could increase the effect, although there is evidence that suggests the converse, meaning that motor activity during speech perception seems emphasized when the speech stimuli are syllables rather than words or sentences (Devlin & Aydelott 2009). Studies dealing with speech perception will be more thoroughly discussed in the context of studies II and III, but it is important to note already that the excitatory effects are observed in regards to both producing and listening to speech.

Sometimes TMS is not even needed in order to observe activation of hand muscles triggered by mouth actions, as was observed by Higginbotham, Isaak and Domingue (2008). They reported that when subjects performed a precision grip, a pointing gesture or a curling gesture, simultaneous activation of the lip muscle orbicularis oris – which is active when producing bilabial stop consonants (e.g. [b]) – is observed. Activation was not observed on facial muscles not involved with labial articulation. The observation of increased activity even without TMS could be taken as an indication of the strength of the connections between hand and mouth motor areas.

In addition to investigating mouth-hand interactions using TMS-EMG techniques, brain imaging studies have also shown similar interactions. For example, in a study utilizing magnetoencephalography (MEG), Salmelin and Sams (2002) studied changes in the brain's mu-rhythm activity, more specifically in its 20 Hz component. The mu-rhythm is an oscillating brain signal, and the 20 Hz component is thought to originate from the pre-central primary motor cortex (Sams & Salmelin, 2002). This oscillation is suppressed by performing movements. Indeed, when participants performed mouth movements (e.g., protruding the lips), Salmelin and Sams (2002) found suppression of the 20 Hz signal in the face area, but most importantly for the current thesis, they also found suppression of the signal in the hand motor area. Interestingly, this suppression was more pronounced in tasks where the participants had to pronounce the vowel [o], touch the upper teeth with the tongue or protrude their lips than in tasks where they had to pronounce actual words. Why this difference existed between verbal and non-verbal tasks can be speculated, but nevertheless, the study of Salmelin and Sams (2002) shows a general linkage between mouth and hand actions.

1.4.1 STUDIES ON GRASPING AND ARTICULATORY GESTURES

Maurizio Gentilucci's group has done a number of behavioural studies regarding the connections between hand and mouth actions. Based on the dual-grasp neuron findings of Rizzolatti et al. (1988), Gentilucci, Benuzzi, Gangitano and Grimaldi (2001) performed a series of behavioural experiments to uncover whether a similar system could exists in humans as well. They measured lip kinematics when participants were instructed to reach and grasp an object with the hand and open their mouth at the same time, and hand kinematics when participants reached and grasped an object with the mouth while opening the hand. The mouth and hand openings were influenced by the size of the grasp opening, that is, the size of the grasped object (Gentilucci et al., 2001). Simultaneous mouth or hand openings were larger, when the grasped objects were larger, compared to if the objects were smaller. Although this is not neurological evidence, these results do support the notion that a similar joint-grasping system as discovered by Rizzolatti et al. (1988) could exist in humans as well.

In another set of experiments, Gentilucci (2003), showed a similar effect in relation to observing grasps. In this study, participants were watching the experimenter perform a reach-and-grasp action and at the same time pronounced a syllable. Both, the size of the mouth opening and the peak amplitude (or intensity) of the vocalisations were larger when the participants observed a larger object being grasped using all fingers on the hand, compared to when a smaller object was grasped with a precision grip using only the thumb and index finger. So, similar effects of hand actions on mouth movements seems to exist for either performing hand actions (Gentilucci et al., 2001) or simply observing them (Gentilucci, 2003). Since the effects are found by simply observing grasp actions, this could imply the mirror system is in place here, in line with the results of Gallese et al., (1996) where the mirror neurons were active both when observing grasping with the hand or mouth.

The effects are not limited to operate from hand to mouth, as Gentilucci and Campione (2011) have shown that finger aperture can be influenced by simultaneous articulations. They had participants articulate different vowels and simultaneously grasp different sized objects. The finger aperture when grasping an object was larger if the mouth was open or an open vowel [ɑ] was pronounced, rather than when the mouth was closed or a closed vowel [i] was pronounced (Gentilucci & Campione, 2011). Thus, it seems that these connections between hand and mouth actions are bi-directional, possibly based on the mirror system and the dual-grasp neurons. Importantly, in the case of mouth actions, the effects are not limited to only grasping, but can be observed also with articulatory gestures. Further, articulating different vowels seem to have different effects on the performed hand grasping.

Besides these general connections, studies have also shown effects of hand actions on the acoustic characteristics of articulations. These characteristics include the intensity (which was already mentioned above), fundamental frequency (f0) and formants (marked F1, F2 etc.). Gentilucci et al. (2001) showed that the intensity of vocalisations of the vowel [a] are higher when, at the same time, a larger object is grasped rather than a smaller one. Similar effects on intensity are observed, when participants just see objects grasped (Gentilucci, Santunione, Roy & Stefanini, 2004; Gentilucci, Campione, Dalla Volta & Bernardis, 2009). The fundamental frequency of speech – responsible for the perceived pitch of vocalisations – determines how high or low the vocalisation is in terms speech melody. The fundamental frequency is higher when participants pronounce [dɑ] while observing a larger object either being grasped or presented alone than when a smaller object is grasped/observed (Gentilucci et al., 2009). Formants are frequency components that have been enhanced by vocal tract resonances and are seen as peaks in the envelope of the sound's frequency spectrum. Formants differentiate vowels from one another (Ladefoged, 2001). Formant labeling starts with F1 being the lowest-frequency formant, F2 the second lowest and so on. Usually only F1 and F2 are enough to differentiate vowels from one another, and vowel diagrams almost always only cover these two dimensions (Ladefoged, 2001). It could be said that these two formants roughly reflect the two-dimensional positioning of the tongue for producing different vowels (Ladefoged, 2001). F1 is generally considered to reflect the openness of the vowel. That is, how open the vocal tract is when producing it (Ladefoged, 2001). This means that for an open vowel, such as [ɑ], the tongue is pushed down, and the mouth is more open. For a closed vowel, such [i], the vocal tract is more narrow, meaning that the tongue is moved fairly high up and the mouth opening is quite narrow. So [ɑ] has a high F1 and [i] has a low F1. In regards to hand actions, F1 is higher when observing a large object being grasped with a power grip compared to when a small object is grasped with a precision grasp (Gentilucci et al., 2009). F2 on the other hand reflects the frontness of the vowel, meaning whether the tongue is positioned more in the front or the back of the mouth during its articulation (Ladefoged, 2001). For example, [i] is a front vowel, meaning that the tongue is placed near the teeth when articulating it, whereas [u] and [ɑ] are a back vowels, meaning that the tongue is placed towards the back of the mouth when articulating them. So F2 is high for [i] whereas it is low for [ɑ] and [u]. For interactions with hand actions, when an apple (a larger fruit) is brought to the mouth and the syllable [ba] is pronounced, the F2 increases compared to if the fruit is a cherry (a smaller fruit) (Gentilucci et al., 2004).

From these findings, it seems then that intensity, pitch and formants F1 and F2 are all higher when the articulations are partnered with grasping larger objects and lower when grasping smaller objects during the articulation. These could be generalised to imply that grasping larger objects affects concurrent vocalisations by making them louder, higher pitched and produced with a more open mouth with the tongue pushed more forward. As such, these effects could be taken as indications about what kind of articulations could be associated with what kind of grasp actions, which is a central part of the next portion, which focuses on a study that was done by our group and that served as the main reference point for this whole thesis.

1.4.2 CONNECTIONS BETWEEN DIFFERENT GRASPS AND ARTICULATIONS

Before getting into the articulations, I want to point out some important differences between our studies and those done by Gentilucci et al. (2001, 2003, 2004, 2009, 2011). One common theme in the studies of Gentilluci's group is the use of large gestures in the experiments. This is of course natural since they wanted to specifically study the kinematics of these actions. Our group wanted to study this from a different methodological viewpoint using grip actions in the absence of the reaching component because our studies focused on measuring reaction times of these actions instead of measuring movement kinematics (Vainio, Schulman, Tiippana & Vainio, 2013). Gentilucci et al. (2001, 2003, 2004, 2009, 2011) studies also focused on the object size and the absolute openings of the mouth and grasp apertures. Instead of the apertures, our group was more interested in associations between specific articulatory gestures and grip types. So, our group decided to study the connections between articulatory gestures and hand actions by using precision and power grip response devices that are constantly held in the hand and only slightly squeezed for the responses (Vainio et al., 2013). The simpler movements reduce the degrees of freedom in the movements and using specific grip devices also forces the participants to make practically exactly the same movements every time. Due to these noteworthy differences between the reach-to-grasp actions and the more static using of grip devices that are constantly held in the hand, I will use the term "grasp" when talking about reach-to-grasp type of movements, that include actually moving the hand to grab onto something. I will use the term "grip" when talking about the small movement of closing the hand perform the action needed to respond with the grip devices, or the very last phase of the reach-to-grasp action.

Moreover, the reason for using power and precision grasps in our studies is that all grasp actions in general can be divided into power or precision grasp (Napier, 1956). Another important aspect for using power and precision grip is the opposing nature of the grips as they were used. Power grip is used to grasp large objects with the whole hand, holding the object against the palm of the hand. Precision grip on the other hand is ideal for grasping small objects by holding the object between the tip of the thumb and the index finger (and/or other fingers).

The importance of this opposing nature of the grips is clearer if the methods used in Vainio et al. (2013) are first explained in more detail. The experiments of these studies utilised a dual-action paradigm, in which two actions were performed simultaneously (grip and utterance). The idea here is that if two actions that are performed simultaneously or in close succession share a common motor planning system, they will be performed faster than if they are coded in different systems (Rosenbaum, 1980). So, if there is specificity in the connections between hand and mouth (e.g. specific grasp is connected to a specific utterance), certain congruent combinations of the two should be faster to perform than other incongruent ones.

Now let us move on to the articulations. Consonant-vowel pairs were used for the vocal responses. The selected syllables were such that in them the articulators are shaped in a way analogous to the grips. The selected syllables were [ti], [pu], [hi], [hu], [pe], [te], [kɑ], [mɑ], [hɑ], [me] and [ke]. These were arranged into pairs [kɑ]-[ti], [mɑ]-[pu], [hɑ]-[hi], [hɑ]-[hu], [me]-[pe] and [ke]-[te] where the first syllable of each pair was predicted to be associated with the power grip second syllable with the precision grip. As mentioned, the open vowel [ɑ] is related to wider finger aperture while grasping and [i] to narrower finger aperture (Gentilucci & Campione, 2011), so it was reasonable to expect [ɑ] to relate to power grip and [i] to precision grip. In addition, the closed rounded vowel [u] was predicted to relate to precision grip since to produce it the lips are protruded, which was linked to fine manual manipulation in chimps (Waters & Fourts, 2002), and the formed mouth aperture is small. These vowel effects were explicitly explored with the [hɑ]-[hi] and [hɑ]-[hu] pairs. In these, the consonant is always the same, and [h] was chosen since as a fricative it was considered a grip-neutral consonant.

The voiceless stop consonant [t] is produced by bringing the tip of the tongue in contact with the alveolar ridge, which can be thought of an analogue to how the tips of the fingers are used to make the precision grip. In contrast, the voiceless stop [k] is made with the back of the tongue coming in contact with the velum. It was speculated that this could be thought of as more of an analogue for the power grip, where the hand is used more wholly to grasp an object. The consonants [p] and [m] are both bilabial, but [p] is a voiceless stop consonant, where the lips are more protruded than in [m], which in contrast is a voiced nasal, where the lip shape is wider and not protruded. Thus, since lip protrusion is associated with precise manual actions (Waters & Fouts, 2002), it was predicted that [p] would be associated more with precision grip and [m] with power grip. These consonant effects were explored in more detail with the [ke]-[te] and [me]-[pe] pairs, where the vowel [e] was the same in both syllables of the pair and as a semi-open vowel, it was hypothesized to be more neutral in its association to the grips.

The experimental procedure in Vainio et al. (2013) was an adaptation of what Tucker and Ellis (2001) used in their study of visuo-motor priming. They found that precision grip reaction times are faster if a viewed object is small (i.e. compatible with precision grip) and that power grip reaction times are faster if the object is large. Like in those studies, in Vainio et al. (2013), participants held both precision grip and power grip devices in one hand. In the experiment, participants were shown first a syllable written in grey colour. Then after a short while, the syllable changed colour and the participants' task was to respond with the grip device that matched the new colour of the syllable, and at the same time pronounce the syllable out loud. It was expected that reaction times would be faster, if the syllable and grip were compatible, as discussed above, compared to if they were not. This is indeed what was found, precision grip reaction times were faster with syllables [ti], [pu], [hi], [hu], [pe] and [te], and power grip reaction times were faster with syllables [kɑ], [mɑ], [hɑ], [me] and [ke] (Vainio et al., 2013). The effect, which I will from now on refer as the articulation-grip congruency (AGC) effect, seemed to be most pronounced with the syllable pair [kɑ]-[ti], where both the consonant and vowel had a grip association. The results of Vainio et al. (2013) thus revealed a systematic association between two motor processes. Although from these results it is impossible to draw detailed conclusions about the background mechanisms, they suggested that there are articulations which mimic precision and power grasps.

1.5 GOALS OF THE CURRENT THESIS

The goal of this thesis was to use the AGC effect to further study the connections between grasping and speech. Below are brief descriptions of the goals of each individual study.

Study I: Study I was the closest adaptation of the original Vainio et al. (2013) study. We wanted to replicate the findings of that study and measure also the vocal reaction times to see whether the AGC effect is also observed in the vocal responses. We expected similar results from the vocal responses as from the manual ones because previous research shows that not only vocal responses can influence manual responses but that this interaction between mouth and hand movements can also operate from hand actions to mouth actions (Gentilucci & Campione, 2011). Another objective was to explore the role of action selection in this effect. We studied whether knowing the required response beforehand removes the effect, or whether the effect persists even if no action selection is needed in the task.

Study II: In Study II, we changed the viewpoint from performed articulations affecting grasps to heard articulations affecting grasps. Research has shown that silent reading (McGuigan 1970) and listening (Fadiga, Craighero, Buccino & Rizzolatti, 2002) of speech is partially processed in the corresponding articulatory representations. Therefore, if articulatory representations indeed interact with the precision and prover grip actions, solely silently reading or hearing syllables would influence responses performed with the grip type that is congruent with the syllable.

Study III: In Study III, we studied whether performing grips could affect speech perception. So, if Study II showed that perceiving speech can influence grasp actions, could this work also in the reverse direction? The motor theory of speech perception (Liberman et al., 1967; Liberman & Mattingly, 1985) suggests that speech perception is based on mapping heard speech to one's own articulatory gestures. That is, speech perception is shaped by first mapping the heard speech sounds to one's own articulatory motor actions (i.e., how one would make those sounds oneself). So, if grasps and articulations share common motor representations, grasp performance could also induce bias to speech perception by producing activity in these shared networks.

Study IV: Study IV was a continuation of Study III. If grasping can bias speech perception, at what level of processing does this influence occur? To this end, we utilized electroencephalography (EEG) and looked at the early (pre-attentive) activity originating from the auditory cortex. We aimed to investigate if the AGC effect can be observed at such an early processing stage.

2 GENERAL METHODS

Participants in all studies were Finnish-speaking adult volunteers. All reported normal or corrected to normal vision, normal hand motor functioning and no known language disorders. All gave written informed consent for participation. The number of participants, gender, handedness and age distributions are presented in Table 1.

Table 1. *Statistics of participants for each experiment of every Study. Experiment 2 & 3 of Study III had the same participants do both experiments during the same session.*

Experiment	N	n of males	left-handed	Age
Study I Experiment 1	17	4	0	24.1 (20–40)
Study I Experiment 2	23	5	1	24.6 (18–29)
Study II Experiment 1	15	2	0	24 (20–27)
Study II Experiment 2	16	2	1	25 (20–31)
Study III Experiment 1	29	0	0	25.3 (19–37)
Study III Experiment 2 & 3	28	7	0	25.1 (19–50)
Study IV	21	2	0	23.6 (20–47)

All studies were approved by the Ethical Review Board in the Humanities and Social and Behavioural Sciences at the University of Helsinki. All experiments were carried out at the Institute of Behavioural Sciences at the University of Helsinki.

All studies except for Experiment 1 of Study III used the same grip devices, presented in Figure 1. The precision grip device was a small cube-shaped device with a micro switch on top. It was held between the thumb and the index finger of the right hand. Responding was done by slightly squeezing the two fingers together, thus activating the switch. The power grip device was cylindrical funnel-like device, held against the palm of the right hand with the remaining three fingers. The power grip device's micro switch was located roughly around the mid part of the device and responding was done by slightly pressing the device against the palm.

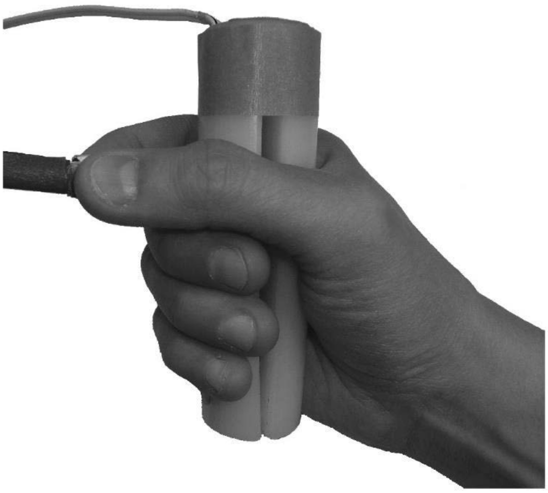

Figure 1 The grip devices used in Studies I, II, IV and Experiments 2 & 3 of Study III, and how they were in the right hand by the participants. Precision grip device is held between the thumb and index finger and is basically just a simple push-button. Reproduced with permission from Vainio et al. (2013).

Grip devices used in Experiment 1 of Study III are presented in Figure 2. The devices were in principle similar to the ones used in other studies but were built to utilize force-resistive sensors to keep track of the force the devices were squeezed with.

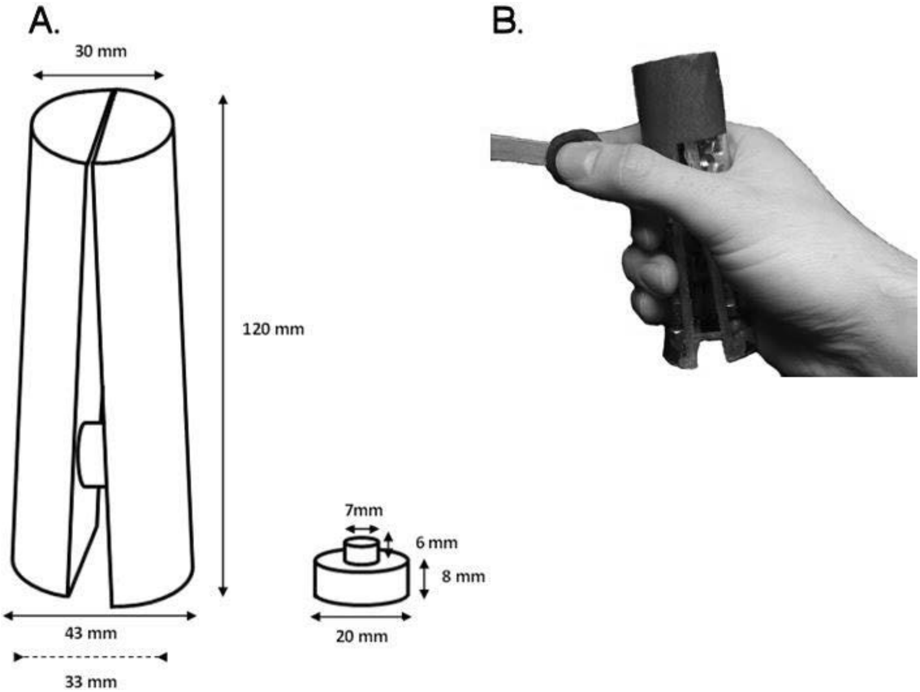

Figure 2 **A**: Schematic drawing of the grip devices used in Experiment 1 of Study III with their associated measurements. **B**: Photo of the devices of how they were held by participants. They were basically the same as the grip devices used in the other experiments except instead of push-buttons they relied on force sensors that were padded with rubber to offer some travel when squeezing the devices. Reproduced with permission from Tiainen et al. (2016).

3 STUDY I – GRIP PLANNING AND VOCALIZATION

Study I consisted of two separate experiments. The primary focus of Experiment 1 was to investigate whether the AGC effect is similarly found with vocalisation responses as it was previously observed with manual responses, since vocalisations were not recorded in the original study (Vainio et al., 2013). The secondary focus, explored in Experiment 2, was to investigate if only preparing a grip response in absence of requirements for selecting the grip response between the two alternatives (i.e., precision or power) is sufficient for observing the effect in vocal responses.

3.1 EXPERIMENT 1 – GRIP EFFECTS ON VOCALISATIONS

As mentioned above, Gentilucci et al. (2001; 2004; 2009; 2011) have found a number of effects of grasp actions on the vocal characteristics of simultaneous articulations. The AGC effect, however, has not been studied in this regard before, and this was the primary aim of Experiment 1. We used the same task as in Vainio et al. (2013) but chose to use only the syllables [ka] and [ti], since they seemed to produce the most robust results in the original study. If the effect is not present in the vocal responses, this would clearly challenge our proposal that the AGC effect reflects an overlap in the planning processes between grasping and articulation. Earlier studies, however, suggest bi-directionality in the hand-mouth connections (e.g. Gentilucci & Campione 2011), which is why we did expect to observe the effect in vocal responses as well. Consequently, we expected that [ka] would be pronounced faster when a power grip is executed than when a precision grip is executed. Conversely, vocal responses of [ti] should be quicker when a precision grip is executed than when a power grip is executed.

Additionally, we could expect to see changes in the vocal characteristics as well. Based on the earlier studies that showed that intensity (Gentilucci et al., 2001), pitch (Gentilucci et al., 2009), F_1 (Gentilucci et al., 2009) and F_2 (Gentilucci et al., 2004) of vocalisations are higher when a large object is grasped and/or power is used to grasp it, we would expect these, then, to be higher for both syllables when a power grip is executed than when a precision grip is executed.

3.1.1 METHODS

Participants sat in front of a computer screen wearing a head-mounted microphone and the two grip devices in their right hand. Participants responded by squeezing one of the devices at the same time as they pronounced a syllable. The syllables in this experiment were [kɑ] and [ti], written on the screen as KA and TI, respectively. On a single trial, one of these syllables was first presented on the screen. After 400 ms the syllable changed colour from grey to either blue or green. When the change happened, the participants' task was to respond as quickly as possible with the appropriate grip device, which were colour coded to match the syllable colour (this coding was balanced across participants), and at the same time pronounce the syllable. The trial structure is illustrated in Figure 3. Each syllable was presented 30 times in each colour (30 x 2 colours x 2 grips = 120 trials).

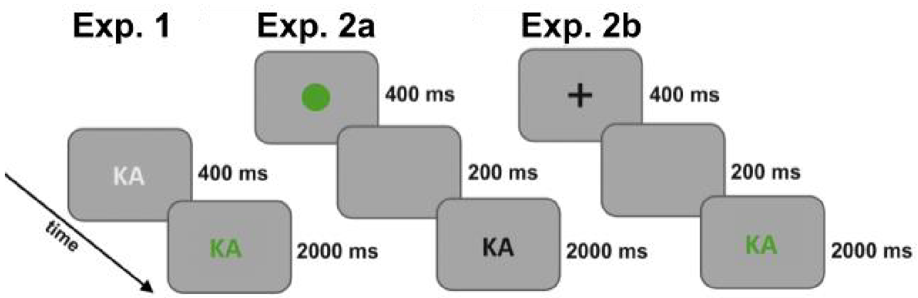

Figure 3 Trial structures for all experiments in Study I. Experiment 1 started with a syllable written in grey that changed colour to either green or blue at 400 ms and the coloured syllable stayed in view for 2000 ms or until a response was made. Experiment 2a started with a coloured circle that indicated the grip response to be made on the trial. It was followed by a short blank screen and then a syllable written in black in lower or upper case letters at which point vocal and/or (depending on the letter case, see text for more details) manual response was to be made according to the colour of the circle at the start of the trial. The syllable stayed in view for 2000 ms or until a response was made. Experiment 2b was structured exactly like 2a, but instead of a coloured circle, a black fixation cross was presented at the beginning and instead of the syllable at the end being written in black, it was written in green or blue, as in Experiment 1.

3.1.2 RESULTS

The results revealed an interaction between grip and syllable for both manual [$F(1,16)=16.73$, $p=.001$, $\eta_p^2=0.51$] and vocal [$F(1,16)=26.48$, $p<.001$, $\eta_p^2=0.62$] responses. The results are presented in Figure 4. In accordance with our previous findings, power grip was executed faster when uttering [kɑ] than when uttering [ti] (411 vs 437 ms, respectively, $p=.013$). Conversely, precision grip was executed faster when uttering [ti] than when uttering [kɑ] (403 vs 437 ms, respectively, $p<.001$). The vocal responses told a similar story: participants were quicker to utter the syllable [kɑ] when a power grip was executed than when a precision grip was executed (477 vs 512 ms, p=.002). Again conversely, [ti] was uttered faster when a precision grip was executed than when a power grip was executed (487 vs 521 ms, p=.002). The effects were partly reflected also in the manual error rates. When a precision grip was required, participants responded with a power grip more often, when the syllable was [kɑ] than when it was [ti] (error rates 3.1% vs 1.7%, Z=-1.96 p=.051). Vocal error rates were too low for analysis.

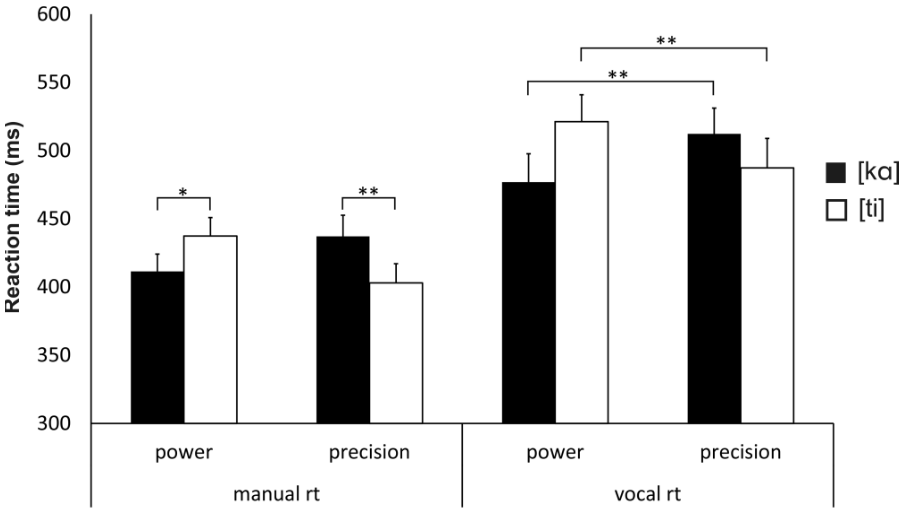

Figure 4 Syllable and grip results for manual and vocal reaction times of Experiment 1. Power grip was performed faster when [kɑ] was pronounced than when [ti] was pronounced. Conversely, precision grip was performed faster when [ti] was pronounced than when [kɑ] was pronounced. For vocalisation reaction times, [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed. [ti] was pronounced quicker when a precision grip was performed. The error bars represent standard errors. Reproduced with permission from Tiainen, Tiippana, Vainio, Komeilipoor & Vainio (2017). * p<.05, ** p<.01.

Contrary to our original hypothesis that grip executions could have specific influences on the vocal characteristics of utterances, there was only one minor finding that [ti] was uttered louder when a power grip was executed (79.0 dB with power grip vs 78.6 dB with precision grip, p<.001).

3.2 EXPERIMENT 2 – ACTION EXECUTION AND ACTION SELECTION

So far, our experiments of the AGC effect have been conducted so that a syllable is first presented, then after a short period it is pronounced at the same time as executing a grip (Vainio et al., 2013). In essence, the vocal response is planned beforehand and then executed together with the precision or power response. In these findings, one of the core factors of the effect (i.e., facilitation of grip responses that are congruent with the vocal response) might be that the participants do not know beforehand which grip they have to perform with the vocalization. That is, the vocal response might bias the response selection processes related to the two response options. Hence, this study explored whether the effect can be observed even when the response selection requirements are removed from the task by informing the participant beforehand which grip they are required to perform. This is the aspect we tackled in Experiment 2 with two separate parts, 2a and 2b.

Until now, I have not talked about the fundamentals that go into any actions that we perform. This is mainly because it is a topic that is largely outside the scope of this thesis, but to understand the intentions of Experiment 2, a short introduction to these matters is necessary. Generally, performing motor actions can be divided into action planning and action execution phases (Glover, 2004). The execution phase is related to the on-line control of movements that mostly operate after the action is already selected. Action planning on the other hand involves higher-level action preparation processing and is responsible for selecting the action that will be executed.

Priming effects in choice reaction time tasks (where a participant must select between two opposing response alternatives) are assumed to be based on cognitive biases that operate within response selection processes (Hommel, 1996). The effect is observed because the task involves a competition between opposite response alternatives. The prime biases this competition so that the response option compatible with the prime receives an advantage, resulting in the facilitation of the compatible response alternative and inhibition of the incompatible response alternative. The effect that perhaps is closest to the AGC effect, at least methodologically, is the size-grip congruency effect, where viewing large objects speeds up power grip responses and viewing smaller objects speeds up precision grip responses (Ellis & Tucker, 2000; Tucker & Ellis, 2001). Importantly for the current context, this size-grip congruency effect vanishes if the grip response is cued before the object presentation (Tucker & Ellis, 2001). Putting this into the context of the above theories, this effect seems to be more based on higher-level cognitive stimulus-response associations that induce bias to the response selection.

On the other hand, following on the lines of these theories, priming effects that are observed in reaction time tasks where the required response is known beforehand can be thought to reflect tight anatomical connections. In other words, these effects are not based on biasing influences of a prime on response selection but rather on some anatomical overlap in processing the prime and response. For example, if participants are told beforehand how to move their finger to respond in a task, seeing a video of the congruent finger moving can still influence the participant's reaction times (Brass, Bekkering & Prinz, 2001). This influence is suggested to be the result of the same motor representations being active for both performing and perceiving the finger movements (Brass et al., 2001).

Since it has been previously suggested the AGC effect to be based on partial overlap of the grasp and articulation motor networks (Vainio et al., 2013), we would expect the effect to persist even if the grip response is known and prepared beforehand. Experiment 2 investigated this by adding a pre-cue for the manual grip response to the experiment protocol of Experiment 2a and then removing all response pre-cues in Experiment 2b.

3.2.1 METHODS

The syllables were the same as in Experiment 1. In Experiment 2a, instead of the syllable written in grey, a coloured circle was presented at the start of the trial. The circle acted as a cue for the required grip response. After the cue there was a blank screen followed by the presentation of the syllable [kɑ] or [ti] written in black, either in upper- or lowercase letters (see Fig. 3 for the trial structure). The upper- and lowercase writing acted as manual go/no-go signals. In the go trials, participants responded as quickly as possible when the syllable was presented on screen with the appropriate grip device (determined by the colour) while simultaneously uttering it. In the no-go trials they only uttered the syllable when it was presented, withholding the prepared manual response. Both colour and go/no-go mappings were balanced across participants.

In Experiment 2b, instead of the initial colour cue, a black fixation cross was presented at the beginning, and the cue for the manual grip appeared simultaneously with presentation of the syllable as the colour of the text (as in Experiment 1), which was again either blue or green (Fig. 3). Otherwise all aspects were identical to Experiment 2a. Colour and go/no-go mappings stayed the same across the two blocks. Order of Experiments 2a and 2b was counterbalanced between participants. Each stimulus combination was presented 30 times in both blocks (30 x 2 colours x 2 grips x 2 blocks = 240 trials in total).

The data were analysed as in Experiment 1 with one exception – the between-subjects factor of response mapping (i.e. the go stimulus for manual response was either upper- or lowercase letters) was added to the ANOVA, making it a mixed-design.

3.2.2 RESULTS

Experiment 2a

The results of Experiment 2a are presented in Figure 5. There was a significant interaction between syllable and grip for the manual [$F(1,21)=16.64$, $p=.001$,

η_p^2=0.44] and vocal [$F(1,21)$=14.79, p=.001, η_p^2=0.41] responses on the go trials as well as on the vocal responses on the no-go trials where there was no manual response [$F(1,21)$=8.29, p=.009, η_p^2=0.28]. On the go trials, precision grip execution was faster when the syllable was [ti] than when it was [kɑ] (463 vs 532 ms, respectively, p<.001), and power grip execution was faster when the syllable was [kɑ] rather than [ti] (473 vs 524 ms, p=.005). For the vocal responses, [kɑ] was uttered more quickly when a power grip was executed than when a precision grip was executed (530 vs 587 ms, p=.001), and [ti] was uttered more quickly when a precision grip was performed than when a power grip was performed (530 vs 578 ms, p=.001). The results for the vocal responses in the no-go trials were similar, when participants were prepared to react with a power grip, they uttered [kɑ] more quickly than when they were prepared to react with a precision grip (510 vs 525 ms, p=.009), whereas participants were quicker to utter [ti] when they were prepared to react with a precision grip rather than with a power grip (511 vs 518 ms, p=.084). Analysis of the manual errors told a similar story: when the required grip was a precision grip, more errors were made when the syllable was [kɑ] than when it was [ti] (error rates 4.0 vs 0.8%, p=.005); when the required grip was a power grip, more errors were made if the syllable was [ti] than when it was [kɑ] (error rates 5.4 vs 1.7%, p=.009).

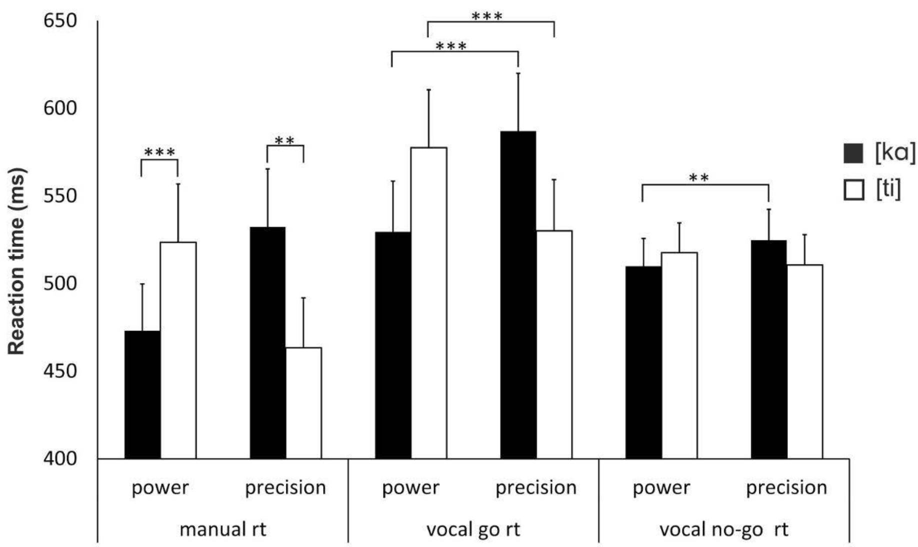

Figure 5 Manual and vocal reaction time results of Experiment 2a. Power grip was executed faster when [kɑ] was pronounced than when [ti] was pronounced. Precision grip was executed faster when [ti] was pronounced than when [kɑ] was pronounced. Both effects were observed even though the manual responses were known beforehand. For Vocal responses, [kɑ] was pronounced faster when a power grip was executed than when a precision grip was performed. [ti] was pronounced faster when a precision grip was Executed. When no manual response was not required (vocal no-go) the effect was similar but slightly smaller. The error bars represent standard errors. Reproduced with permission from Tiainen et al. (2017). ** p<.01, *** p<.001.

There were also secondary findings of the interaction between syllable and response mapping and between grip and response mapping. When participants were reacting to lowercase letters, they responded manually and vocally faster to the syllable [ti], and manually also when the grip was a precision grip. When participants reacted to uppercase letters, vocal responses were faster when the grip response was a power grip.

Experiment 2b

Results of Experiment 2b are presented in Figure 6. The syllable-grip interaction was significant for both manual [$F(1,21)=63.27$, $p<.001$, $\eta_p^2=0.75$] and vocal [$F(1,21)=49.95$, $p<.001$, $\eta_p^2=0.70$] responses in the go trials, and approaching significance in the no-go trials [$F(1,21)=3.97$, $p=.059$, $\eta_p^2=0.16$]. In the go trials, the precision grip was performed faster when [ti] was pronounced than when [kɑ] was pronounced (595 vs 652 ms, p<.001), and the power grip was performed faster when [kɑ] was pronounced than when [ti] was pronounced (604 vs 664 ms, p<.001). Similarly, [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed (637 vs 677 ms, p<.001), and [ti] was pronounced faster when a precision grip was performed than when a power grip was performed (638 vs 686 ms, p<.001). The manual errors revealed that when the required grip was a precision grip, more errors were made when the syllable was [kɑ] than when it was [ti] (error rates 5.2 vs 1.4%, p=.001); when the required grip was a power grip, more errors were made if the syllable was [ti] compared to [kɑ] (error rates 8.2 vs 1.6%, p<.001).

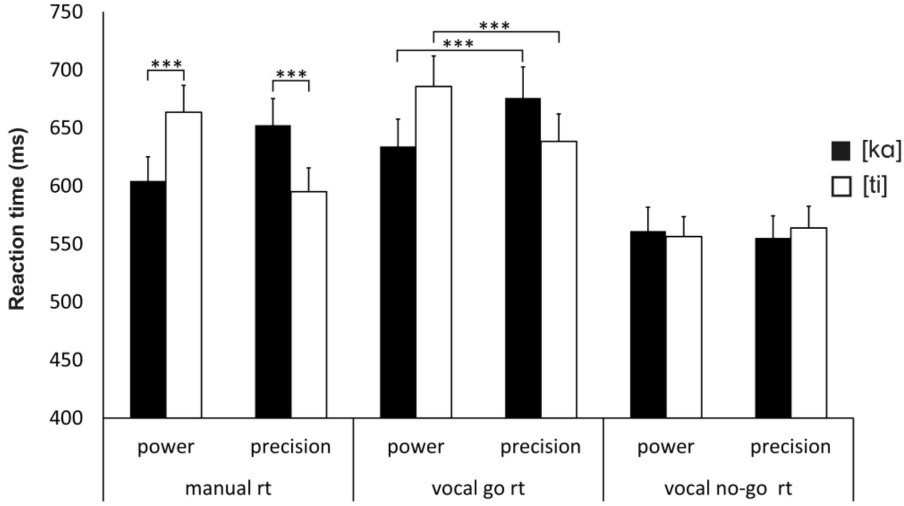

Figure 6 Manual and vocal reaction times results of Experiment 2b. Power grip was executed faster when [kɑ] was pronounced than when [ti] was pronounced, and precision grip was executed faster when [ti] was pronounced than when [kɑ] was pronounced. Vocally, [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed, and [ti] was pronounced faster when a precision grip was performed. When the manual response was inhibited, the interaction was marginally significant and reversed. The error bars represent standard errors. Reproduced with permission from Tiainen et al. (2017). ** p<.01, *** p<.001.

4 STUDY II – INFLUENCE OF HEARD AND READ SYLLABLES ON GRIP EXECUTION

Study I showed that the AGC effect can be observed in vocalisations and that just preparing a grip can affect vocalisations. Study II, then, took this idea forward by asking whether perceiving the syllable by reading or hearing it could influence grip execution.

Overt and covert articulations are generally thought to involve the same neural mechanisms (Palmer et al., 2001). This is supported by the fact that listening to speech and covert articulations are related to increased mouth muscle activity (McGuigan 1970; Fadiga et al., 2002). Also, listening to speech sounds that require tongue movements increases tongue muscle activity (Fadiga et al., 2002). These findings are in line with the motor theory of speech perception, which suggests that speech perception is based on interpreting observed speech in one's own articulatory representation system (Liberman et al., 1967). Furthermore, in relation to the AGC effect, Devlin and Aydelott (2009) showed that motor involvement in speech perception is emphasized when the stimuli are syllables rather than words. Therefore, like in Study I, where just preparing a grip was sufficient to produce the AGC effect, the overt articulation, or even intent to articulate the syllable might not be necessary. The Stroop task is a well-known example of how people automatically process phonemic aspects of viewed words (e.g., Bakan & Alperson 1967; Dennis & Newstead 1981). In the task, participants are required to list out loud the colours in which a list of words is written in. The interference comes from the fact the words whose colour the participants are listing, are themselves names of colours. For example, the word "GREEN" written in blue colour, so the participant should respond with the word "blue". The automatic processing of the viewed words however impairs the performance significantly, compared to if the words are not related to colours or if the word is congruent with the actual colour (e.g., word "BLUE" written in blue colour). Perhaps this kind of automatic processing of perceived syllables is all that is needed to observe the AGC effect as well. In other words, presentation of a syllable (for example auditorily) during a task where a participant is supposed to execute precision and power grips, the automatic processing of the syllable could facilitate or impair the grip execution depending on whether the syllable and grip are congruent or incongruent.

4.1 EXPERIMENT 1 – INFLUENCE OF READING SYLLABLES ON GRIP EXECUTION

Experiment 1 was a lot like the original Vainio et al. (2013) study, but instead of overtly pronouncing the presented syllable, they had to only read it silently. In Study I Experiment 2a, the participants were always prepared to pronounce the syllable, but had to withhold doing so, depending on the onscreen cue of a trial. Here on the other hand, the participants never had the intention to pronounce the syllable. Thus, observing the effect in this experiment would again provide evidence for the joined network of grasps and articulatory gestures, and imply that Stroop-like automatic processing of the phonetic aspects of the viewed word is enough to trigger the AGC effect.

4.1.1 METHODS

During the experiment, participants sat in front of a monitor with the two response devices in their right hand. The stimuli consisted of the syllables [ti], [kɑ], [py] and [mo] (written as TI, KA, PY and MO, respectively). The procedure was like Experiment 1 of Study I, in that the syllable was first displayed in grey and then it changed colour, at which point the participants responded with the correct device as quickly as possible (see Fig. 7). The critical difference was that participants were not allowed to pronounce the syllable aloud. In addition, participants were instructed to not respond if the syllable was either PY or MO. This way, the participants were forced to read the syllable. In total, the experiment consisted of 216 trials of which 96 were no-go trials and 120 were go trials.

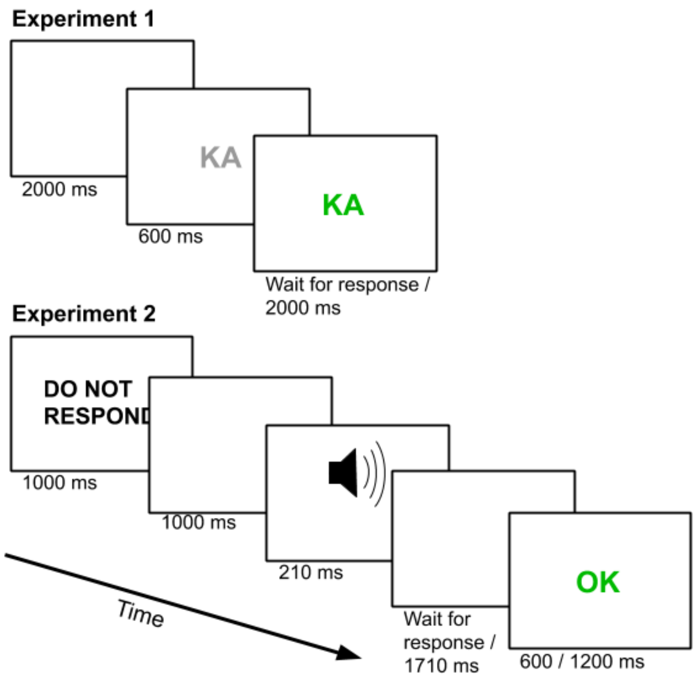

Figure 7 Trial structures for Experiment 1 and 2 in Study II. Experiment 1 started with a blank screen for 2000 ms, followed by a syllable written in grey that the participants had to read. After 600 ms the syllable changed colour to either green or blue and stayed in view for 2000 ms or until a grip response was made according to the colour of the syllable. Experiment 2 started either with the text “ÄLÄ VASTAA” (“do not respond” in Finnish) for 1000 ms followed by a blank screen for 1000 ms, or with just a blank screen for 2000 ms. Then a syllable was played through headphones, after which the participants were required to answer with a grip device according to whether the syllable was presented in low or high pitch. Finally a feedback text was presented to indicate whether the grip response was correct or not.

4.1.2 RESULTS

The interaction between syllable and grip, presented in Figure 8, was significant [$F(1,14)=14.19$, $p=.002$, $\eta_p^2=0.503$]. Precision grip responses were faster when the syllable was [ti] (412 ms) rather than [kɑ] (441 ms). Power grip responses were faster when the syllable was [kɑ] (440 ms) rather than [ti] (455 ms). This interaction demonstrates that the AGC effect can be observed even when the syllable was read silently.

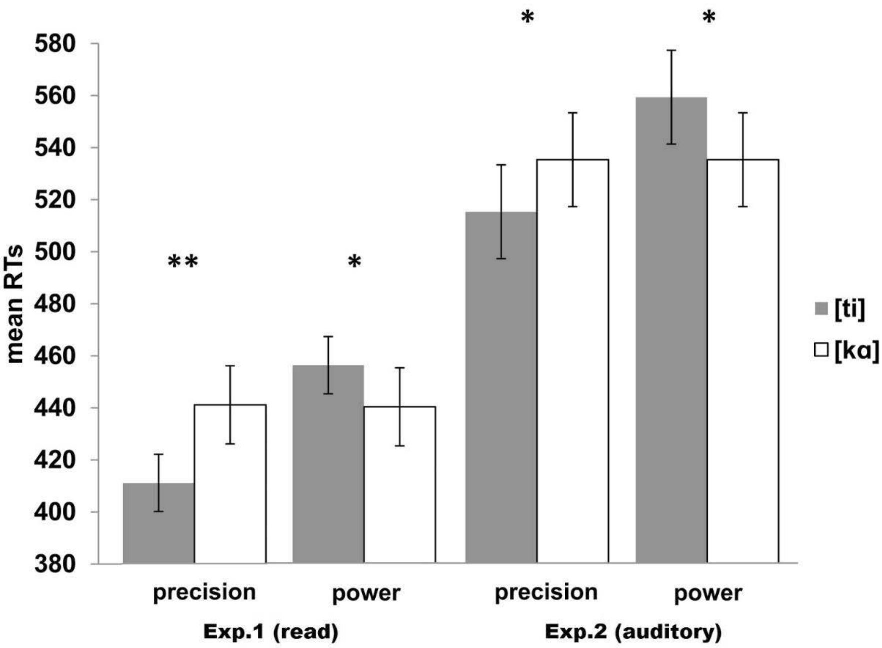

Figure 8 Reaction time results for Experiments 1 and 2 of Study II. In both experiments precision grip was executed faster if the presented syllable was [ti] than if the syllable was [kɑ]. Vice versa for power grip, it was executed faster if the presented syllable was [kɑ] rather than [ti]. The error bars represent standard errors. Reproduced with permission from Vainio, Tiainen, Tiippana & Vainio (2014). * p<.05, ** p<.01.

4.2 EXPERIMENT 2 – INFLUENCE OF LISTENING TO SYLLABLES ON GRIP EXECUTION

Like in Experiment 1, there were no overt articulations in Experiment 2. But instead of silently reading the syllables, the syllables were played to the participants through headphones. It has been shown that heard or seen speech increases the responses of mouth muscles compared to non-speech in TMS stimulation (Watkins, Strafella & Paus, 2003). Similarly, functional magnetic resonance imaging (fMRI) has shown that motor areas that are active during speech production are similarly active also during hearing of the same speech sounds (Pulvermüller et al. 2006). Additionally, this kind of motor activity during speech perception seems to be underlined when the speech stimuli are syllables rather than full words or sentences (Devlin & Aydelott 2009). Hence, if there is indeed an overlap between grasping and articulatory gestures, one might predict that solely hearing a syllables would influence manual grip responses congruent with the syllable. Hence, based on these notions, we expected to observe the AGC effect on the grip responses when the syllables are only heard.

4.2.1 METHODS

The equipment was the same as in Experiment 1 but the stimuli were auditory. The stimuli consisted of the syllables [ti] and [kɑ]. Both syllables were pronounced by a female. The voice parameters of the stimuli were manipulated so that the lengths syllables were identical (210 ms) and the intensities were equalized. High-pitched and low-pitched versions of the sounds were made by raising and lowering the pitch of the syllables. The participants' task was to respond to the pitch of the stimulus and ignore the actual syllable. In one block of the experiment, the participants were instructed to respond to the high voice with the precision grip and to the low voice with the power grip and in another block the other way around. The order of the blocks was balanced between participants. The trial began with a blank screen or the text "ÄLÄ VASTAA" ("DO NOT RESPOND" in Finnish) that indicated a no-go trial. The purpose of the no-go trials was to make sure that participants kept their eyes on the monitor instead of their responding hand. The blank screen/text was followed by the auditory stimulus. Illustration of the trial structure is presented in Figure 7. In total, the experiment consisted of 240 trials, 30 trials for each stimulus combination.

4.2.2 RESULTS

The interaction between syllable and grip was significant [$F(1,14)=10.85$, $p=.005$, $\eta_p^2=0.437$]. Precision grip responses were faster when the syllable was [ti] (516 ms) rather than [kɑ] (535 ms, $p=0.020$) and power grip responses were faster when the syllable was [kɑ] (536 ms) rather than [ti] (559 ms, $p=.040$). This interaction, shown in Figure 8, shows that the AGC effect can be triggered by just hearing a syllable that is congruent with the grip.

There was also a secondary finding for the pitch-compatibility (high-pitch and precision grip, low pitch and power grip), responses were faster in pitch-compatible (515 ms) than in pitch-incompatible (558 ms) conditions, $F(1,14)=7.68$, $p=.015$, $\eta_p^2=0.354$.

5 STUDY III – GRIP INFLUENCE ON SPEECH CATEGORIZATION

In Study II we found that listening to syllables influences grip execution, which led us to contemplate, whether this connection could work in the reverse direction as well. That is, could grip execution or planning influence the categorization of syllables in a similar manner?

So-called perception-action theories suggest that perceptual and motor processes operate in, at least partially, shared system (e.g. Pulvermüller & Fadiga, 2010). In this system perception can influence motor processes, but interestingly, motor actions can also influence perceptual processes. In support of this latter direction in terms of grasping actions, studies have shown that preparing a grasp response makes the processing of a viewed object faster if the viewed object matches the prepared grasp (Craighero, Fadiga, Rizzolatti, & Umiltà, 1999; Symes, Tucker, Ellis, Vainio, & Ottoboni, 2008). For example, when participants plan a grasp before introducing changes in a presented scene (a so-called change blindness task), the participants are more likely to be aware of the changes in objects that are congruent with the planned grasp (e.g. power grip planned and changing object is an apple) (Symes et al., 2008). So basically, planning the grasp action improves the perception of congruent objects, which reduces the blindness to changes in them.

So, there is evidence that grasping actions can influence perceptual processing on congruent objects, what about speech? If we think about the considerations of the motor theory of speech perception (Liberman et al., 1967), it would seem likely that similar effects would exist regarding speech as well. Indeed, providing TMS over the language production areas can influence syllable categorization (D'Ausilio, Bufalari, Salmas & Fadiga, 2012; Meister, Wilson, Deblieck, Wu & Iacoboni, 2007; Möttönen & Watkins, 2009). For example, providing repetitive TMS (rTMS) over the lip area to temporarily disrupt its functioning also impairs the categorical perception of syllables that require lip movements to produce (e.g. [ba]–[da]) but not of those that do not require the lips (e.g. [ka]–[ga]) (Möttönen & Watkins, 2009). Similar results have been obtained in behavioural studies by having participants perform overt silent articulation (Sams, Möttönen & Sihvonen, 2005; Mochida et al., 2013; Sato, Troille, Ménard, Cathiard & Gracco, 2013). When participants are instructed to articulate a syllable silently in synchrony of a heard syllable, the articulation influences the categorization of the heard syllable (Sams et al., 2005). For example, when pronouncing [ka], auditorily presented syllable [pa] is categorized more often as [ka] than in a control condition, where there is no simultaneous articulation.

Considering this evidence for how grasps and articulations can influence visual and auditory perceptual processing, we could expect to find that categorization of syllables congruent with the precision or power grip would be systematically influenced by performing these grips. This is what Study III explored in a set of three experiments.

5.1 EXPERIMENT 1

In Experiment 1 we had participants prepare and execute either a power or precision grip while we presented them a syllable and then had the participants report which syllable was presented. We chose to use syllables [ke] and [te] as the stimuli instead of [kɑ] and [ti] that were used in the previous studies, since using different vowels in the syllables would make the syllables too easy to recognize. The potential influence of the ceiling effect on the measured responses was also controlled by adding noise to the stimuli.

With the above considerations, we expected to observe the AGC effect in relation to categorizing perceived syllables, meaning that we expected more [ke] responses when a power grip is prepared and executed, and conversely more [te] responses when a precision grip is prepared and executed. We expected to see this effect in relation to all three different modality conditions: visual, auditory and audiovisual speech. Since seen speech increases the responses of mouth muscles in TMS stimulation similarly to auditory speech (Watkins et al., 2003), we could expect similar results here across all modalities.

5.1.1 METHODS

In the experiment, the participants sat in front of a monitor, holding the power and precision grip response devices in their right hand, listening to the auditory stimuli through headphones. The speech stimuli used were utterances of syllables [ke] and [te] by a female speaker. The utterances were embedded in pink noise (more power at the lower frequencies) to prevent a ceiling effect in recognition. The utterances were presented audiovisually (talking head with audio), auditorily (audio only) or visually (talking head without audio).

The trial started with a coloured circle which specified the grip to be used on the trial. The colour mapping was balanced between participants. The participants had to first squeeze the appropriate device lightly to prepare the correct grip and to start the trial. After the trial started, the syllable was presented in one of the three modalities, and the participants were told to hold the light pressure during the utterance and respond with a quick squeeze when the utterance ends. After the grip response, they reported which syllable was presented with a left-hand keypress. The options were “k” or “t”, as the vowel was irrelevant in the task. There were also no-grip trials where no grip response was performed and the trial started with a black fixation cross instead of the circle. Figure 9 shows the trial structure with a sample of the grip-force data. Each stimulus combination was presented 20 times.

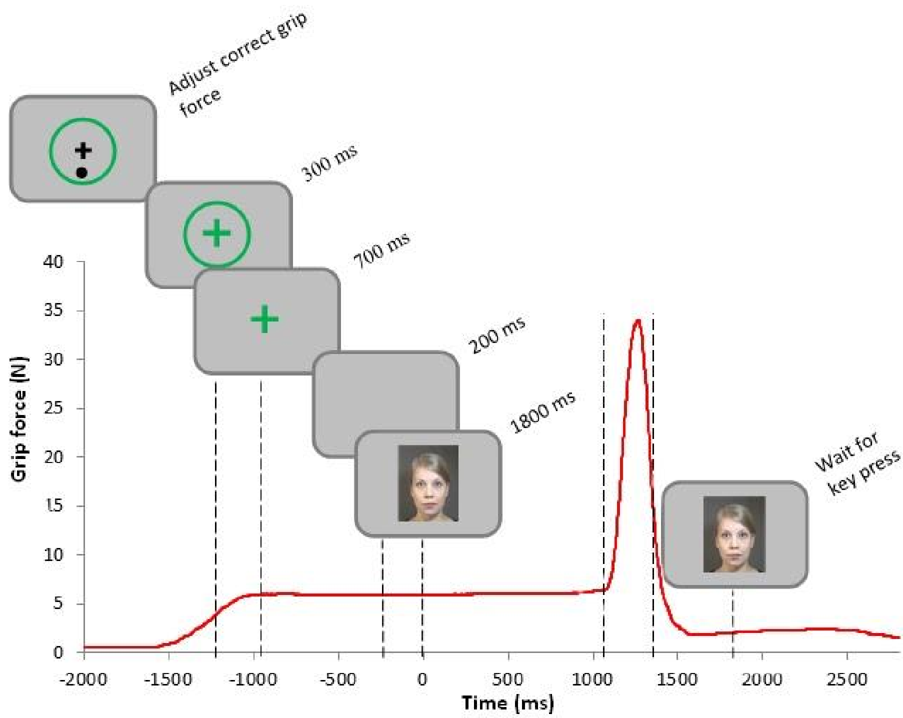

Figure 9 The trial structure of Experiment 1 of Study III. The red line represents the changes in grip-force on a single trial. The zero point on the time axis is the stimulus onset. Before the zero point is the time period for readying the correct grip (see text for details). The dashed line at 1100 ms marks the end of the auditory stimulus and the dashed line at 1300 ms marks the ends of the visual speech articulation movements. At around 1800 ms the visual stimulus ended completely and participants were to give their response for which syllable was presented. After selecting the response with a key press there was a 1000 ms interval before the next trial started. Reproduced with permission from Tiainen et al. (2016).

5.1.2 RESULTS

The interaction between syllable and grip was significant [$F(2,56)=21.31$, $p<.001$, $\eta_p^2=0.43$] (Figure 10). The syllable [ke] was categorized correctly more often on power grip trials (.82) than on precision grip (.74, $p<.001$) or no grip trials (.77, $p<.001$). The syllable [te] was categorized correctly more often on precision grip (.77) and no-grip trials (.75) than on power grip trials (.71, $p=.001$, $p=.028$). The interaction was similar in visual, auditory and audio-visual trials, so only the aggregate results are presented here.

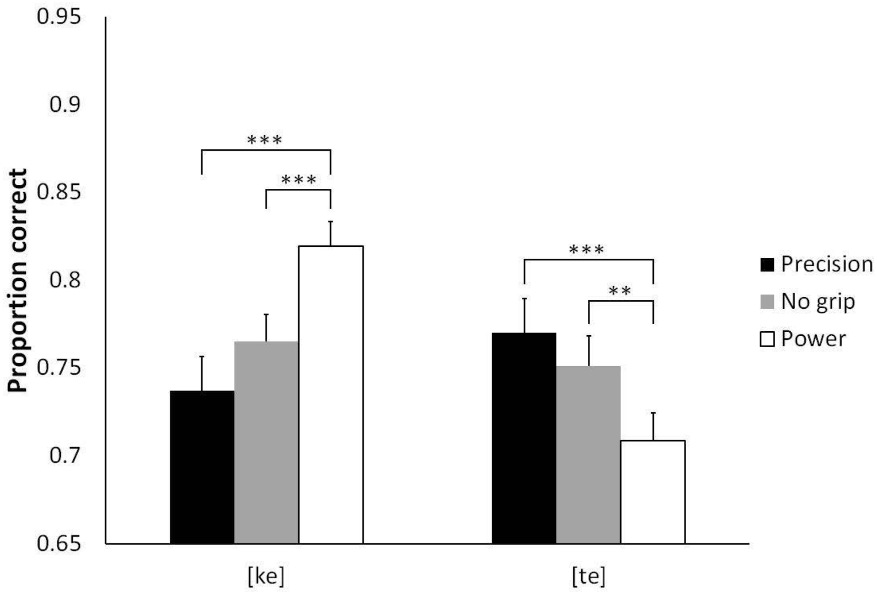

Figure 10 The proportions of correct responses between the two syllables. The responses are averaged over all modalities (auditory, visual and audiovisual). The proportion of [ke] responses was larger when the grip response was a power grip than then when there was no grip executed or the grip was a precision grip. The proportion of [te] responses was larger when the grip response was a precision grip than when there was no grip response or the grip response was a precision grip. The error bars represent standard errors. Reproduced with permission from Tiainen et al. (2016). ** = p < .01, *** = p < .001.

5.2 EXPERIMENTS 2 & 3 – GRIP AND SPEECH PERCEPTION WITH AND WITHOUT RESPONSE PRE-CUE

Experiment 1 did produce the AGC in regards to syllable categorization. However, this does not necessarily mean that the actual perception was influenced, as the grasps could have just modulated the decision-making process of reporting the syllable. In fact, even the claims of most studies about mouth movements affecting speech perception have been challenged, other mechanisms that do not involve perception modulation could explain those results (Hickok, 2010). This suggests that the results of Experiment 1 might reflect a response bias, or a tendency to systematically favour a certain response over the other in a specific condition. That is, it is possible that an action performed in conjunction with the categorization task influences the response selection process responsible for reporting the recently perceived syllable. This is why we decided to follow up with two additional experiments that were designed to address the response bias question.

Experiment 2 was a replication of Experiment 1 in a simplified form. The overt preparation of a grip response was not required and instead there was just a pre-cue at the beginning of the trial and a go signal at the end of the syllable. We also included only audiovisual trials, which allowed us to increase the number of trials. We wanted to have more trials so that we could utilize signal detection theory (SDT, e.g. Green & Swets 1966; Macmillan & Creelman, 1991) in the analysis. SDT allowed us to characterize each participant's performance on two parameters, discriminability (*d'*) and criterion (*c*). The ability to discriminate between the two stimuli, [ke] and [te], is reflected by the d'. So, an increase in *d'* caused by grip performance would indicate an enhancement in the perception of the syllables. The c parameter on the other hand reflects the categorization boundary between the two syllables, so basically the tendency to favor one response over the other. This basic idea of SDT is visualized in Figure 11. SDT has actually been recently used to address the critique towards the speech production-perception effects (Smalle, Rogers & Möttönen, 2014). The study showed a *d'* effect so that the difference between syllables "ba" and "da" was less salient after disrupting the mouth motor representations via TMS. They did not find significant changes in *c*, which suggests that the effect is indeed perceptual.

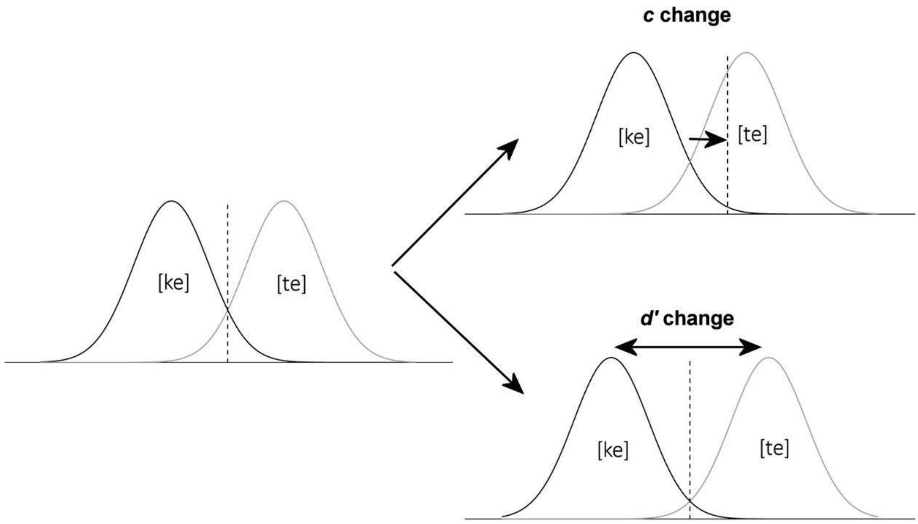

Figure 11 Theoretical illustrations of changes in *d'* (distance between the two distributions) and *c* (the dashed line that represents the decision boundary between a [ke] and [te] perception). On the left side a situation where *d'* = 3 and *c* = 0 (no bias). Values to the left of the *c*-line represent situations when the person would report hearing [ke] and values to the right situations when the person would report hearing [te]. In the top right is a situation where the *c* = 1. In this case almost all [ke] responses would be correctly recognized, but there would be a significant increase in false categorizations of [te] syllables as being [ke]. In the bottom right on the other hand there is a situation where *d'* = 4, meaning that the actual discriminability of the two syllables has increased from the left hand situation. Now both syllables are almost always recognized correctly, without any increase of false reports. Reproduced with permission from Tiainen et al. (2016).

However, a criterion effect may actually arise from the perceptual level as well (e.g., Witt, Taylor, Sugovic, & Wixted, 2015). The underlying signal distributions are what define the criterion's location. So, if the underlying distributions would shift equally, but the criterion would stay the same, it would still manifest as a criterion effect without a change in the discriminability. In practice this would mean that in a power grip context, both of the syllables [ke] and [te] become perceptually more power grip-like, (i.e. more [ke] -like), while the criterion stays the same. In Experiment 3 the response bias question was addressed through the task procedure rather than through analytical methods. The procedure was the same as in Experiment 2, except that the initial response cue was removed. So, the only cue for the grip response was the go signal's colour. This means that a grip could not be prepared before the syllable presentation. Additionally, if the grip would have an effect on the categorization, it would have to result from an influence on post-perceptual processes since syllable presentation had already ended when the grip was selected and executed.

5.2.1 METHODS

Experiments 2 and 3 used the same stimuli as Experiment 1, but only the audiovisual stimuli. Experiments 2 and 3 were conducted in the same session with Experiment 3 performed first, since it was considerably shorter than Experiment 2. This way we attempted to avoid any effects of tiredness or getting overly used to the stimuli that might have followed if the the longer Experiment 2 was done first. In Experiment 2 a coloured fixation cross (green, blue or black) acted as the cue for the grip to be executed in the trial (power, precision or no grip). After the fixation cross, the speech stimulus was presented. At the end of the utterance, when the speaker's mouth closed, a transparent coloured circle appeared on top of the face. This was the go-signal for the grip response and was the same colour as cue. Colour mapping was balanced between participants, except for the no-grip cue, which was always black. After the grip, participants reported which syllable was presented ([ke] or [te]). Each stimulus condition was presented 60 times.

Experiment 3 was identical in structure to Experiment 2, except that the fixation cross at the beginning was always black. Thus, the participants could not prepare the grip response beforehand. Figure 12 illustrates the structure of both experiments. Each stimulus condition was presented 20 times (as in Experiment 1).

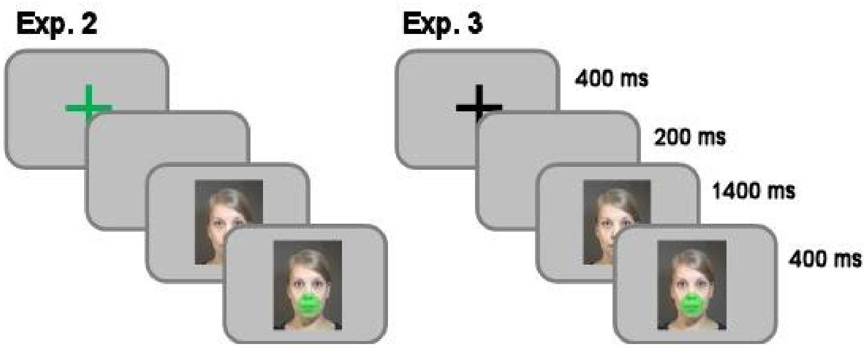

Figure 12 Trial structure of Experiments 2 and 3 of Study III. The fixation cross at the beginning of the trial acted as a pre-cue in Experiment 2, the colour representing the required grip response. The fixation cross was presented for 400 ms and followed by a blank screen for 200 ms. After this the stimulus presentation began. At 1400 ms after the stimulus onset, when the talker had closed her mouth, a coloured partly transparent circle appeared on top of the speaker's mouth that acted as the go-signal for the grip response. After this the participants reported the syllable that was presented, as in Experiment 1. The trial structures of the two experiments were identical except for the fact that the fixation cross at the start of the trial in Experiment 3 was always black, meaning that there was no information about the grip response before the coloured circle was presented after the syllable had ended. Reproduced with permission from Tiainen et al. (2016).

5.2.2 RESULTS

Experiment 2

The interaction between syllable and grip was significant [$F(2,54)=15.23$, $p<.001$, $\eta_p^2=0.36$]. [ke] was categorized correctly more often on power grip (.83) and on no-grip (.81) trials than on precision grip trials (.77, $p=.003$ and $p=.037$). In contrast, [te] was categorized correctly more often on precision grip (.89) and no-grip (.87) trials than on power grip trials (.83, $p<.001$ and $p=.005$). These results are presented in Figure 13 A.

Signal detection analysis results are presented in Figure 13 B. This analysis revealed only a criterion effect [$F(2,50)=14.93$, $p<.001$, $\eta_p^2=0.37$], favouring more [te] (vs. [ke]) responses with precision grip (c=-0.23) than with power grip (c=0.04), or when there was no grip (c=-0.07) (Fig 6B).

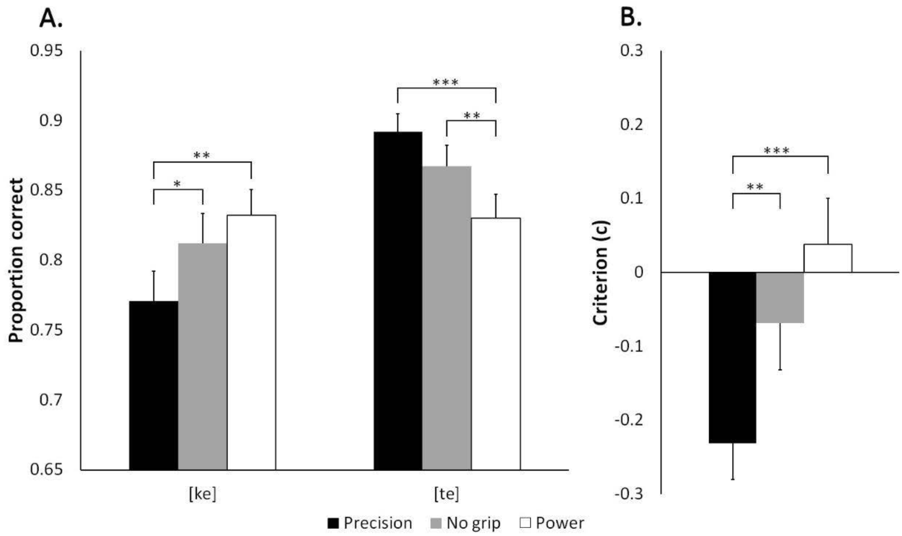

Figure 13 Results for Study III Experiment 2. **A**: Proportions of correct responses for the two syllables. When the syllable was [ke], participants made more correct categorizations, if the prepared grip response was power grip than if no grip or precision grip was prepared. When the syllable was [te], participants made more correct categorizations if the prepared grip was a precision grip rather than if no grip or a power grip was prepared. **B**: The criterion (*c*) values of different grips from the STD analysis. Positive values indicate a criterion that favors [ke] response and negative values a criterion that favors [te] response. Zero-point is the optimal criterion with no bias. Power grip was associated with a positive and precision grip with a negative criterion value. There were no significant effects for *d'*. The error bars represent standard errors. Reproduced with permission from Tiainen et al. (2016). * = p < .05, ** = p < .01, *** = p < .001.

Experiment 3

The interaction between syllable and grip was significant [$F(2,54)=4.86$, $p=.011$, $\eta_p^2=0.15$]. There were more correct [te] responses when a precision grip was executed as opposed to no grip. Other comparisons were non-significant, but the interaction appeared similar to that observed in Experiment 2 (Fig. 14), and a combined analysis of Experiments 2 and 3 revealed no difference between the two interaction effects (syllable × grip × experiment interaction p=.349).

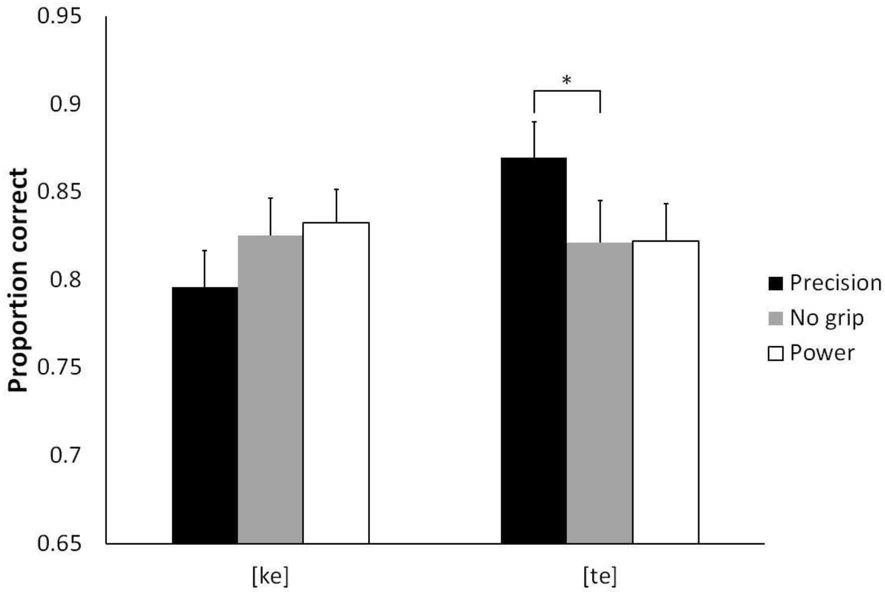

Figure 14 The proportions of correct responses in Experiment 3, where all grip-related information was given post-syllable presentation. When the syllable was [ke], there was no significant difference between the different grip conditions. When the syllable was [te], participants made more correct categorizations if the executed grip was a precision grip than if there was no grip executed. The error bars represent standard errors. Reproduced with permission from Tiainen et al. (2016). * = p < .05.

6 STUDY IV – GRIP INFLUENCE ON SYLLABLE PROCESSING AT THE NEURAL LEVEL

Study III revealed that the AGC effect could also influence speech categorization. However, the results of Experiments 2 and 3 of Study III suggested that this influence was based on response bias induced by the grip. However, the difference in robustness between the results of Experiment 2 and 3 might suggest that there is something more than pure response bias involved. In Study IV we wanted to study the early processing influences of grips by measuring early brain responses. If grips could modulate early, pre-attentive, brain responses to presented stimuli, then it is possible that the categorization effect could also be partly truly perceptually based.

6.1 EEG AND THE MISMATCH NEGATIVITY

In Study IV, we used EEG to measure participants' brain activity during the experiment. We utilized the well-known auditory event-related potential (ERP) component, the mismatch negativity (MMN, e.g. Näätänen, Paavilainen, Rinne & Alho, 2007). Traditionally the MMN is recorded in what is called an oddball paradigm. In this paradigm, the participant is presented with a continuous stream of auditory stimuli, called standard and deviant stimuli. Standard stimuli comprise the majority of the auditory stimuli, we used a percentage of 85% standard stimuli. Among the stream of standard stimuli, occasionally a deviant stimulus is presented, in our case the percentage of deviant stimuli was 15%. This sudden occurrence of a deviant stimulus in a stream of otherwise constant standard stimuli causes an automatic response in the EEG signal. The MMN can be observed from the difference curve between the standard and deviant stimulus. The actual MMN component is observed as negativity in the difference curve that peaks usually around 100-200 ms after stimulus onset. The MMN represents a mismatch between the predicted regularity of the signal and actual heard signal (Winkler, 2007). The origin of the signal has been traced to the auditory cortex (Alho, 1995). Since it is observed even if no attention is paid to the audio signal, it is considered to be a pre-attentive component related to early, relatively automatic stages of auditory processing (Näätänen et al., 2007). Although the MMN was originally observed using simple tones, it is also observable using speech stimuli (Näätänen et al., 2007). An important discovery in respect to this Study is the fact that the MMN can be modulated by visual information (e.g., Sams et al., 1991; Colin et al., 2002; Möttönen, Krause, Tiippana & Sams, 2002). This means that other information can influence auditory processing at a pre-attentive level. Maybe it could be possible to influence MMN also with motor actions if there is a strong connection between the sounds and movements, as there seems to be in the AGC effect.

We used the same [ke] and [te] utterances as in Study III in an oddball design with two separate blocks, where one syllable acted as the standard stimulus and the other one as the deviant. The task itself was relatively simple. In it, the participants had to respond with the grip devices according to onscreen colour cues that determined which grip to execute. The sounds themselves could be completely ignored, as is usually the case in the MMN paradigm. Since both [ke] and [te] appeared as the deviants in separate blocks, we had conditions that were physically identical, just the context changed. For example, a condition where a power grip was executed, while a standard sound [ke] was presented versus a condition where a power grip was executed and a deviant sound [ke] was presented. If we could observe a modulation in the MMN caused by the grips it would support the hypothesis that grasp actions can influence speech processing also at the perceptual level. In addition, this would be the first time that actions are shown to be capable of modulating MMN patterns.

6.2 METHODS

In the experiment the participants sat in front of a computer screen holding the grip devices in their right hand. The experiment used an oddball paradigm, where one stimulus was the standard (P=.85), and the other the deviant (P=.15). The experiment was split into two blocks where in one the standard was [ke] and the deviant was [te] and vice versa in the other one. The syllables were the same as in Study III, but without the noise masking. The order of the blocks was balanced across participants. The participants' task was to ignore the audio and simply to respond to onscreen coloured circles with the appropriate grip devices. The circles were timed to appear 100 ms before the syllable started, so that the actual grip processing could start around the same time as the syllable processing. Illustration of the trial structure can be seen in Figure 15. There were six types of trials: standard audio with no grip (2150 trials), standard audio with a power grip (300 trials), standard audio with a precision grip (300 trials), deviant audio with no grip (150 trials), deviant audio with a power grip (150 trials) and deviant audio with a precision grip (150 trials). In total, one block had 3000 trials.

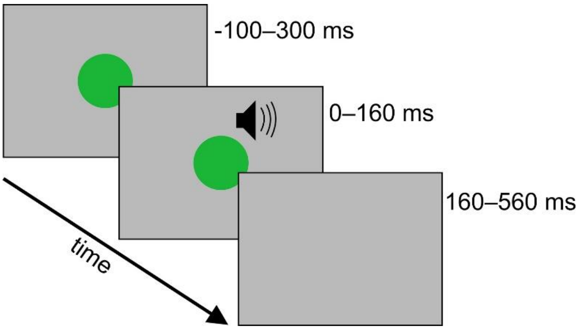

Figure 15 Trial structure for the experiment of Study IV. The zero-point for the trial was the syllable start and the syllable lasted for 160 ms. The participants did not need to pay attention to the syllable. If the trial required a grip response, a coloured circle was presented on screen from 100 ms before the syllable start to 300 ms after the stimulus started. The colour signalled the grip to be made. The participants had until 560 ms after to execute the grip response. Thus there was a 400 ms gap between auditory stimulus presentations. When no grip response was needed, the screen remained blank. Reproduced with permission from Tiainen, Tiippana, Paavilainen, Vainio & Vainio (2017).

The EEG was recorded with a 64-electrode cap with five additional electrodes: at the right and left mastoids, tip of the nose, below the left eye and at the outer canthus of the left eye. Vertical electro-oculogram (EOG) was recorded with the electrode below the eye and horizontal EOG with the electrode on the side of the eye. The sample rate used was 512 Hz, which resulted in a dynamic range of DC–204Hz (2/5 of the sampling rate).

The nose electrode was used as the reference in the analysis. The EEG data were first band-pass filtered (0.5–40 Hz). Channels with noisy signal were removed for each participant individually based on visual inspection. Then, independent component analysis (ICA), and more specifically the runica-algorithm, was used to identify and remove the signal components that contributed most to the eye blink signal. The components to remove were chosen manually by comparing them to the EOG signal. Two components at most were removed for any single participant. After this, the signal was split into EEG epochs for each trial, from 100 ms before to 500 ms after the auditory stimulus onset. The 100 ms prestimulus time was used as the baseline for amplitude measurements. Only trials where the grip response was correct were used for ERP averaging. This lead to rejecting 2.4% of trials. Trials with voltage changes higher than ±100 µV were also removed. In total 15.7% of trials were rejected. One participant's data had to be completely rejected due to corrupted data. Remaining participants had at least 73 trials for each trial type (average on deviant-grip trials was 128).

We calculated difference curves by subtracting the ERPs for standard trials from the corresponding deviant trials. For example, from the standard [te] with precision grip, we subtracted the deviant [te] with precision grip. This means that physically identical trials were used in the subtraction (same grip and syllable), with only a context difference of the sound, acting as the deviant or the standard stimulus. The most notable negativity peaks (within typical MMN occurrence window of 100-200 ms) and their durations were used as the time windows for the statistical analyses. On the basis of this selection, the time window for [ke] was 100–200 ms and 150–250 for [te]. Mean amplitudes were calculated for electrodes F3, F1, Fz, FC3, FC1, FCz, C3, C1 and Cz. Left-side electrodes were selected as the MMN in speech-related tasks originates mainly from the left hemisphere (Näätänen et al., 1997) and because some participants had poor signal on the right-side electrodes.

Repeated-measures ANOVA (2 × 2 × 3 × 3 design) was performed for the selected electrodes with the factors deviant syllable ([ke] and [te]), grip (power and precision), electrode row (i.e., the frontal-central axis: F, FC and C) and electrode column (i.e., the left-midline axis: 3, 1 and z). Only the trials where grip responses were executed were compared, because the modulation in MMN by the two grips was the main point of interest. Greenhouse-Geisser correction was used when the sphericity assumption was violated. Partial eta-squared served as the effect size estimate. Pairwise comparisons were Bonferroni-corrected.

6.3 RESULTS

The difference curves (Fig. 16, bottom) showed an MMN on each trial type, peaking during a 150–250 ms time window after stimulus onset in the deviant [ke] block and during 100–200 ms time window in the deviant [te] block. The distribution of the amplitudes (Fig. 16, top) showed a typical MMN pattern: the largest amplitudes were recorded at the fronto-central areas over the left hemisphere. The MMNs to deviant [ke] were larger than those to deviant [te].

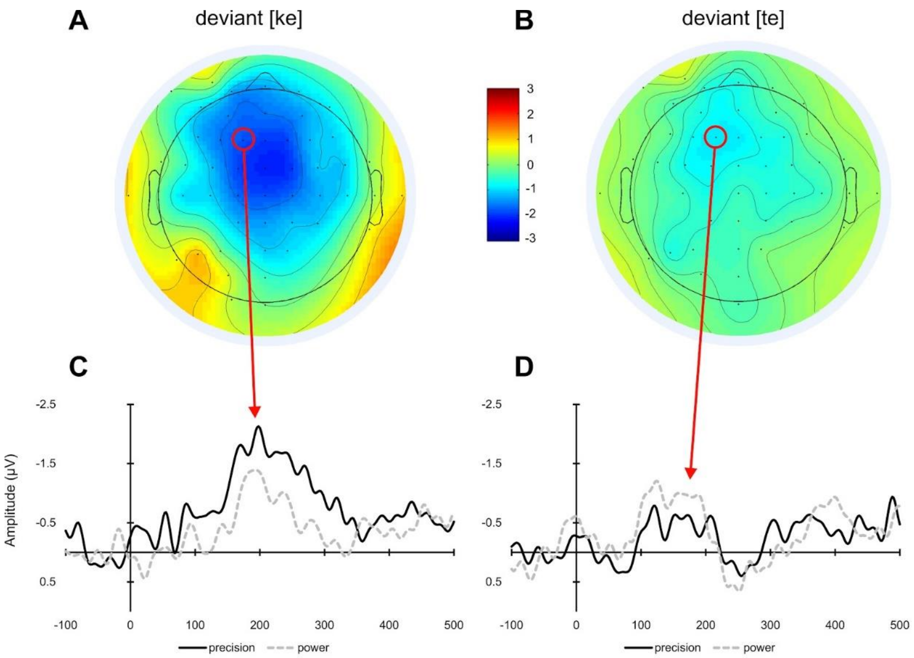

Figure 16 **A:** Isopotential map for deviant [ke] precision grip trials for 180–220 ms time window. **B:** Isopotential map for deviant [te] power grip trials for 140–180 ms time window. **C:** Example ERP grand average difference curves on electrode F1 from the deviant [ke] block. **D:** Example ERP grand average difference curves on electrode F1 from the deviant [te] block. Reproduced with permission from Tiainen, Tiippana, Paavilainen et al. (2017).

Analysis of the frontal left side electrodes (F3, F1 and Fz; see Table 2 for the MMN amplitudes) revealed a significant syllable-grip interaction [$F(1,19)=6.29$, $p=.021$, $\eta_p^2=.25$, power=.66]. The MMN was larger with [ke] for precision grip than for power grip (-1.662 µV vs. -0.864 µV, $p=.042$) and there was a trend that the MMN was larger for power than precision grip with [te] (-0.882 µV vs. -0.439 µV, $p=.132$). These results are also illustrated in Figure 17. The difference between [ke] and [te] was significant for precision grip ($p=.037$), the MMN was larger for [ke] than [te]. Difference between syllables was not significant for power grip ($p=.972$).

Table 2. *Average MMN amplitudes in µV (standard deviations in parentheses) for the frontal electrodes for each grip trial.*

	deviant [ke] (150-250 ms)		deviant [te] (100-200 ms)	
Electrode	Precision	Power	Precision	Power
F3	-1.52 (1.37)	-0.69 (1.44)	-0.39 (1.54)	-0.80 (1.55)
F1	-1.68 (1.58)	-0.98 (1.39)	-0.49 (1.50)	-0.94 (1.70)
Fz	-1.79 (1.67)	-0.93 (1.72)	-0.44 (1.44)	-0.91 (1.61)

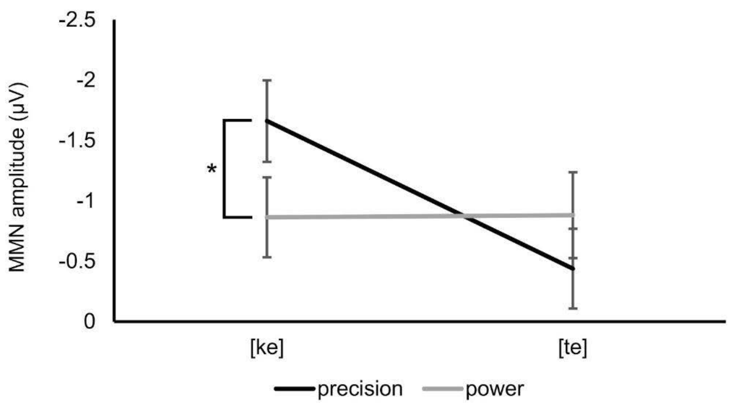

Figure 17 Mean amplitudes for the electrodes F3, F1 and Fz of the MMN signal. When [ke] was the deviant stimulus, the MMN was larger when a precision grip was executed than when a precision grip was executed. A converse trend was observed when [te] was the deviant, but this difference was not significant. The error bars represent standard errors. * = p < .05.

7 DISCUSSION

In this thesis, I have presented four studies that show the robustness of the AGC effect and expand our understanding of how it operates. The effect is observed in vocal responses both when a simultaneous grip is executed and when it is only prepared (Study I) and in manual responses when the syllables are only heard or read silently (Study II), or when the manual response is known beforehand (Study I). The results of Study III also showed that grip preparation can influence speech categorization processes. The idea that grip preparation and execution can influence the processing of speech stimuli congruent with the grip type at an early perceptual processing state was supported by the results of Study IV where brain responses related to early auditory processing were modified by grip performance. In the following I will first discuss the results of each study individually and then all of them as a whole.

7.1 STUDY I

In Study I, the AGC effect was observed in vocal reaction times for the first time. For the vocal responses, [ti] utterances were quicker when a precision grip rather than a power grip was executed and [kɑ] utterances were quicker when a power grip rather than a precision grip was executed. In addition, the previously observed AGC effect in manual responses (Vainio et al., 2013) was also observed. For the manual responses, power grip responses were faster when the utterance was [kɑ] rather than [ti], whereas precision grip reaction times were faster when the utterance was [ti] rather than [kɑ]. These results support the bidirectionality between mouth and hand actions shown in previous studies (e.g., Gentilucci et al., 2001; Gentilucci & Campione, 2011). An important extension to those studies here is that these results provide evidence about the specificity of the connections at the syllable/grip level, as the studies by Gentilucci and his colleagues (e.g., 2001, 2011) measured how opening/closing the mouth influences aperture of the precision grip.

The AGC effect on manual reaction times was observed even when the manual response was known a priori and only needed to be executed in synchrony with the utterance (Experiment 2a). The AGC effect was also similarly observed in vocal reaction times when the vocal response was known a priori (Experiment 1). That is, the response selection is not required in the task in order to observe the effect. As mentioned in the introduction of Study I, an effect observed in these kinds of situations would suggest tight overlap between the components, i.e. the articulations and grips. This is in line with our proposal that the AGC effect is based on shared mechanisms between articulatory and manual gestures (Vainio et al., 2013), and not on some learned higher-level (e.g., semantic) associations between the two. Furthermore, the vocal AGC effect was observed even when the manual response was just prepared and not carried out (Experiment 2a). This suggests that the effect operates at the level of action planning as the execution of the action is not required to observe the effect.

We also found an effect of grip influencing the intensity of [ti] vocalisations, where power grip was associated with louder vocalisations. Previously Gentilucci et al. (2009, 2004) have reported that when participants watch objects being grasped, increasing object size is associated with increasing intensity of vocalisations. These results could be seen as being somewhat in line with each other, since larger objects are associated with power grip in general (Tucker & Ellis, 2001). However, since this result was observed only in Experiment 1, and by and large vocal characteristics were not affected by the grip responses, there is little basis to argue for these influences in our experiments.

Given the numerous findings concerning interactions between articulations and manual gestures by Gentilucci et al. (2001; 2009; 2004), the lack of influence of manual actions on vocal characteristics here was unexpected. The reason for this difference in results could lie in the responses used. The manual responses in our experiments required very minimal hand movements, since the participants were always holding the response devices in their hand. In contrast, Gentilucci et al. (e.g. 2004) have used larger movements, usually involving moving the whole arm in a reach-to-grasp manner. Thus, it is possible that larger and more sustained movements might be needed in order to observe the influence of manual movements on the vocal characteristics.

7.2 STUDY II

In Study II, the AGC was observed in manual responses both when participants silently read syllables or listened to them. That is, power grip reaction times were faster when silently reading (Experiment 1) or hearing (Experiment 2) the syllable [kɑ] rather than [ti] and precision grip reaction times were faster when hearing or silently reading the syllable [ti] rather than [kɑ]. These results are evidence for the robustness of the AGC effect, but also require the AGC effect to be somewhat redefined. As overt articulation was not needed for the effect, the activation of the articulation's mental representation – which can be achieved also by listening to the syllable or covert articulation – is enough to produce the AGC effect. Covert articulation has indeed been shown to be an implicit, automatic part of word recognition (e.g., Eiter & Inhoff, 2008). Thus, it should be emphasised that the A in the AGC effect includes also covert articulations. These results complement nicely those of

Study I, where the vocal AGC effect was observed when the grip was only prepared, whereas here the manual effect was observed when the syllables were only read silently. In addition to this, they provide complementary evidence that the manual AGC effect is also not just a by-product of the synchrony between manual and vocal responses discussed above, since here the manual effect was observed when the articulations were not physically executed. Furthermore, the results of Experiment 2 suggest that the effect influences particularly the grasp planning. This is because there was a significant gap after the auditory stimulus before the grip response. All in all, the results of Study II together with the results of Study I support the view that the AGC effect is based on the tight and interconnected network between motor representations of specific articulatory gestures and grip actions. When one component of this network (e.g., articulatory gesture for the consonant [t]) is activated due to any reason (e.g., it is heard, pronounced or read silently), the manual counterpart of this component (e.g., motor representation of the precision grip) is automatically activated facilitating responses that utilize this manual representation.

Gentilucci, Benuzzi, Bertolani, Daprati and Gangitano (2000) found that when grasping objects with “GRANDE” (large) written on them, the finger aperture at the initial phase of the movement was greater than when grasping an object with the word “PICCOLO” (small) written on it. The same effect was observed by Glover and Dixon (2002) using English words “LARGE” and “SMALL” printed on the objects. These effects were quite naturally explained as higher level processing of the word and its semantics being added to the grasp planning and thus influencing the hand movements. However, the results of Experiment 1 suggest that this does not need to be the case. Grasp actions could be influenced by the phonetic properties of the written words in the absence of any direct semantic primes. Indeed, the above effects could partially (but not necessarily fully) be explained just by the phonetic properties of the words, since “GRANDE” contains [a], which is associated with power grip, and “PICCOLO” contains [i], which is associated with precision grip.

One might also assume that the results of Experiment 2 were, to some extent, in line with the suggestions of the motor theory of speech perception that speech perception involves activation of the articulatory representations that would be involved in producing the heard speech sounds (e.g., Liberman et al., 1967). Our results extend this suggestion so that listening to speech also creates activity in the shared network of articulations and grasp actions, resulting in the observed effect that listening to syllables affects grip execution based on the associations between the phonetic properties of the sound and the grip. The finding that simply listening to speech during TMS stimulation of hand areas increases hand muscle activity (Flöel et al., 2003) is in line with these views and results. From the idea of speech perception influencing grasp actions, it is natural to transition to Study III and the idea of grasp actions influencing speech perception.

7.3 STUDY III

While Study II suggested that speech perception can influence grasp actions, in Study III we found that grasp actions can influence speech categorization. Participants reported hearing more [ke] utterances when executing a power grip and more [te] utterances when executing a precision grip. This was observed despite the fact that the grip and speech stimuli were not directly associated with each other. The effect was similar with auditory, visual and audio-visual speech stimuli (Experiment 1). Further experiments showed that the effect was at least partly based on biases of selecting one of the two syllables for response caused by the grip performance before selecting the syllable response. By utilizing SDT analysis in Experiment 2, we discovered a criterion effect of grip, which suggested that the grips influenced the subjective criterion of how the participants categorized the two syllables. That is, executing a power grip made the participants more likely to categorize the syllable as [ke] and executing a precision grip made them more likely to categorize the syllables as [te]. A possible discriminability effect would have meant that executing the grips would make the participants more accurate in general in discriminating between the two syllables, which was not observed.

Results of Experiment 3 further clarified the picture, as the grip effect on categorization was observed even when all grip related information was given after the syllable was presented but the grip was executed before the categorization response was given. This means that there was practically no possibility that the grip could have influenced the perceptual processing of the speech stimulus. Instead, the grip influenced the selection of the syllable after the actual perceptual processing. In other words, this supports the view that the effects observed in this study were at least partly based on biasing influence of grip performance on syllable categorization and/or selection rather than influence of grip performance on syllable perception.

We assume that the results of Study III are based on interconnected network between articulatory gestures and grip representations in which executing, for example, a precision grip could cause activation in the involved networks, including the articulatory component for the consonant [t]. This in turn increases activation in the [t] component of the network. Now, because of the increased activity in the [t] component, the participant would be more biased to select the syllable [te] for the response regardless of what they actually perceived.

In the spatial-numerical association of response codes, or SNARC, effect, a left-right key press following the presentation of a digit is biased by the digit's numerical value (Daar & Pratt, 2008). In general, low digits facilitate left responses and high digits facilitate right responses. Another example of the SNARC effect is that when participants are asked to freely select and pronounce a number (in the range of 1-40), they execute more spontaneous eye movements that are directed right and up before pronouncing a high number, and eye movements directed left and down before pronouncing a low number (Loetscher, Bockisch, Nicholls, & Brugger, 2010). This is proposed to reflect a biasing effect of the preceding action (i.e. eye movement) on the upcoming action selection (i.e. choosing the number to pronounce) (Loetscher et al., 2010). Comparing our results to these, Study III provides the first evidence that action-induced bias can be observed in this kind of language processing context, and more specifically in the context of manual grasping actions and speech.

Differentiating between perception, action and decision-making processes is difficult (Cisek & Kalaska, 2010), and there could be multiple overlapping effects in play here: the biases of grip performance on selecting the syllable and the influence of grip performance on perceiving syllables. The effect size was largest in Experiment 1 and smallest in Experiment 3, while Experiment 2 was between the two. Hence, it is possible that in Experiment 1 in which the grip performance was planned and executed prior to the syllable categorization task, the relatively large effect reflects both effects. Consequently, the effect was particularly large. In contrast, in Experiment 3 in which the grip performance was planned and executed after offset of the categorized syllable, the effect purely reflects only the grip-related biases on syllable selection. consequently, the effect was particularly small. That is, the more grip representations are active, the more notable the effects are on speech perception. All in all, the results suggest that grasp actions can act as context cues, much like the segments of the actual speech that can help deduce unheard parts of speech in noisy situations, even for already heard speech. The actual levels of processing (perception/decision making, or both) that this operates cannot be concluded from these results, but this is something that was the aim of Study IV.

7.4 STUDY IV

In Study IV, we used EEG and created a novel MMN paradigm to further explore the possible influence of executing power and precision grips on speech processing, based on the results of Study III. We did indeed find an interaction between syllable and grip. When a precision grip was executed and the deviant syllable was [ke], the MMN was larger than when a power grip was executed. There was a reverse trend when the deviant was [te] (MMN with power grip larger than with precision grip), although this was not statistically significant. The MMN is regarded as a pre-attentive signal originating from the auditory cortex (Alho, 1995). So, although the task itself was not a perceptual one, if the manual grasps can influence processing at this stage, it is likely that it could influence following perceptual processes.

We did not have a specific hypothesis about how the two grips would modulate the MMN amplitude, as this was a very explorative study. However, we speculated that as precision grip is associated with [te], in the deviant [ke] block precision grip execution enhances the processing of the standard [te] sound, since the participant is already prepared to hear [te], but then when the deviant [ke] is played, the result is an increased mismatch between the two syllables, and thus a larger MMN response is observed. In the deviant [te] block, the setup is reversed, and the power grip enhances processing of the standard [ke], and mismatch between [ke] and [te] is increased. Even though this latter difference between standard [ke] and deviant [te] was not significant, a significant difference from zero in the deviant [te] block was observed only when a power grip was executed. This indicates that the power grip did enhance the difference between the syllables.

In general, the MMN was weaker in the deviant [te] block, which is a likely reason for why the difference between grips was not significant. This means that there were some unknown features in the stimuli that made [te] less effective deviant than [ke]. Thus, it is possible that the difference between grips would have been significant, had the MMN been larger in the deviant [te] block. The difference between blocks could be approached from another perspective as well. The difference between the syllables was related to precision grip, and not to power grip. Power grip might thus not have influenced the syllable processing enough to produce an effect in the MMN whereas the precision grip did. With the response devices both held in one hand, the precision grip is performed rather normally, but due to the precision device occupying the thumb and the index finger, the power grip needs to be executed somewhat unnaturally with only the middle, ring and little finger. This mismatch should be addressed in further studies.

These results should be taken as indicative rather than confirmatory. The effect was rather small, and the paradigm was not without problems. The effects were not symmetrical, and we struggled to get a good MMN-response for the deviant [te]. The grip executions also induced a lot of additional noise to the EEG signal, which might have masked some of the effects. Altering the experimental procedure and stimuli could result in better data and clearer, more reliable effects. Even though a major reason for Study IV was to clarify the results of Study III, the fact remains that Study IV was not a direct measure of perceptual changes. Therefore, the possibility of perceptual changes induced by grips still needs further clarification.

Although these results are far from conclusive, together with the ones from Study III, they suggest that activating the grip representations induces implicit decision bias when categorizing syllables but also inflicts changes in processing already in the auditory system. More specifically, the results from Study IV suggest that there is a possibility that manual grips can influence speech processing also at the perceptual stage, actually even before the conscious processing. This is all in line with the motor theory of speech perception and the idea that speech perception is achieved by simulating the heard speech in the listener's own motor system (Liberman et al., 1967), but extends this view with the idea that it is not just the mouth motor representations, but also hand motor representations that are activated.

7.5 GENERAL DISCUSSION

Taken together, all the studies reported in this thesis speak for the robustness of the AGC effect. It was observed in manual responses with and without overt articulation; in vocal responses with and without actual grip execution, regardless whether the articulation was self-produced or heard. It was even shown to affect how participants categorized perceived syllables. In addition to these studies, we have replicated the original effect (Vainio et al., 2013) with Czech (Tiainen, Lukavský et al., 2017) and English (Tiainen, Felisberti et al., 2016) speakers. This indicates that the effect is not language-specific. We have even reported similar effects related to other movements, namely vertical hand movements (Vainio, Tiainen, Tiippana, Komeilipoor, & Vainio, 2015; Tiainen Lukavský et al., 2017). Indeed, all these point to the conclusion that there exist strong connections between specific articulatory and hand movements. Although this is well in line with the gestural theories of language evolution, these results cannot be taken as confirmatory evidence for these theories.

Previous studies have also shown connections between speech and hand movements (e.g., Gentilucci et al., 2001; Gentilucci 2003; Gentilucci et al., 2004, Gentilucci et al., 2009; Gentilucci & Campiano 2011). However, these previously presented connections operate at a relatively coarse level, connecting, for example, increased grasp aperture with increased lip opening when grasping a large object in comparison to small object. Our studies expanded on those results by studying connections between specific manual and articulatory gestures, namely between syllables (such as [kɑ] and [ti]) and manual grips (i.e., power and precision grip). That is, the results highlight the suggestion that the connections operate at the level of strictly specific grips and articulations. The results of this thesis suggest that these connections between manual and articulatory gestures are robust in that they are observed in many different tasks influencing even early processing of auditory speech stimulus.

In general, the AGC effect seem to be stronger in relation to precision grip than to power grip. For example, Study IV showed the effect so that precision grip was the one that increased the size of the MMN in the deviant [ke] block. Similarly, in Experiment 3 of Study III, the only significant difference was the increase of correct [te] responses when performing a precision grip. Expanding on the ideas of MacNeilage's (1998) frame/content theory, maybe the power grip is not so much related to specific articulations, but refers more to the "frame" of speech, i.e. the general opening and closing of the mouth and the rhythm of speech. This could be supported by the findings of Rizzolatti et al. (1988), where they did not find any neurons that were specific to grasping with a power grip and with the mouth. This could indicate a more general connection between the power grip and articulatory representations. The observations of Waters and Fouts (2002) also support this, as they observed mouth movements on chimpanzees concurrently with fine motor movements, such as precision grasps, but not with gross motor movements such as power grasps. This would explain why the AGC effect appears to be stronger in relation to precision grip and the articulations associated with it.

Given that I talked a lot about the evolution of language in the Introduction, it is relevant to also discuss how our findings relate to those theories, specifically the gestural theories of language evolution. Unfortunately, it is practically impossible to get any concrete evidence of how language evolved. All that can be stated is that our results are generally in line with the gestural theories of language evolution (e.g., Arbib, 2005; Hewes, 1973, Gentilucci & Corballis, 2006). Regardless of the evolutionary aspect, the results here strongly support the idea that hand and mouth motor actions form a partly shared network, where the overlapping connections can be quite specific (i.e. connecting a specific grip to a specific articulatory gesture). This again is in line with the observation of the double-grasp neurons that react both to grasps performed with the hand and with the mouth (Rizzolatti et al., 1988). I would further speculate that the strength of activations in one domain are correlated with the effects observed in the other, as was somewhat seen in Studies I and III: simply preparing a grip produces an effect on vocalisations, but actually executing the grip makes that effect stronger.

I would also argue that the AGC effect seems to originate at a "lower" (e.g., less semantic) level than, for example, the size-grip effect, where power grip is associated with large objects and precision grip with small objects (e.g., Tucker & Ellis, 2001). The size-grip effect disappears if the grip response is known a priori, whereas the AGC effect persists even in this situation (Experiment 2 of Study I). Compared to earlier findings of the connections between grasp actions and articulation (e.g., Gentilucci et al., 2001), our results accentuate the importance of the types of the grasps and articulations. The effects of the grasps change markedly depending on the grasp type.

I started this thesis saying that gestures are an integral part of our daily lives. However, basically this whole thesis has focused on small components of speech (i.e., syllables) and manual gestures (i.e., grip types) that are detached from the focus of our daily communication. Differentiating between noisy [ke] and [te] syllables while performing different grips is probably not something most of us do on a daily basis. This is of course because our research aimed to show the underlying mechanisms behind interactions between manual actions and speech. But are there any practical implications? It could be proposed that these results could provide insights for treating speech disorders. I would speculate that with the knowledge that precision grip is associated with the consonant [t], difficulties in a child's learning of the consonant [t] could be helped by performing precision grips when practicing the consonant. Furthermore, considering the evidence of hand-mouth connections in infants/children and influence of precise manual skills in the development of speech (e.g., Fogel & Hannan, 1985; Iverson & Thelen, 1999; Nelson et al., 2014), deeper knowledge of these connections and of their specificity could prove helpful in the future in understanding the development of language skills of children. In fact, Vainio (2019) has recently suggested, partly based on the results of the studies presented here, that learning of specific vocalisations in childhood (such as [k] and [t]) might be facilitated by the learning of specific manual gestures (such as power and precision grasping).

There were also a couple interesting secondary findings in the studies reported here. I will discuss these very shortly, since they are not in the focus of this thesis. For the grip responses, there were interactions also with letter size (Study I) and pitch (Study II), and for articulations with letter size (Study I). Power grip responses were associated with lower pitch and syllables written with capital letters; precision grip responses were associated with higher pitch and syllables written in lowercase letters. [kɑ] articulations were also related to capital letters and [ti] to lowercase. One might assume that these effects are based on similar sensorimotor processes that are also responsible for the size-grip effect (Tucker & Ellis, 2001). Given that small objects typically tend to resonate at higher frequencies than large objects (see Spence, 2011), it is possible, for example, that the precision grip is associated with the higher pitch because it can be semantically conceptualized as a small concept. Nevertheless, these indicate a more complex web of connections for grips and articulations that seems to exist independent from the grip-articulation connections.

In general, the studies would have benefitted from a wider sample pool. There was a clear overrepresentation of female participants. Although we did not consider it likely that there would be any difference in the results between men and women, because of the lack of male participants this cannot be stated for sure. Also, practically all participants were university students. By acquiring participants from outside the university we could have increased the representativeness of the samples. Thus, future studies should aim to recruit participants also from outside of the university and aim for a better balance between male and female participants.

8 CONCLUSIONS

Specific connections between articulation of syllables and different grasp actions were found in four tightly connected studies, such that the syllables [kɑ] and [ke] were associated with a power grip action, whereas the syllables [ti] and [te] were associated with a precision grip action. The connections were very robust. They were observed both in vocal and manual responses, even when the grip was only prepared and not executed, and when the syllable was only heard or read silently. There were even effects of grip actions on syllable categorization. Further neural evidence from EEG suggested that the interaction effects between syllables and grips might originate from a pre-attentional processing stage. I propose these results to reflect that hand motor representations and articulatory representations form a partly shared network, where activity from one domain can induce activity in the other. This is in line with theories that language has evolved from or alongside a manual communication system.

9 REFERENCES

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing, 16(1),* 38–51.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28(2),* 105–124.

Babkin, P. S. (1958). Early postnatal establishment of reflex activity in man. *Fiziologicheskii Zhurnal SSSR Imeni IM Sechenova, 44(10),* 922–927.

Bakan, P., & Alperson, B. (1967). Pronounceability, attensity, and interference in the Color-Word Test. *The American Journal of Psychology, 80,* 416–420.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica, 106(1),* 3–22.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry, 36(1),* 1–22.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33,* 269–298.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113(4),* 495–506.

Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: a motor-visual attentional effect. Journal of experimental psychology: *Human Perception and Performance, 25(6)*, 1673–1692.

Daar, M., & Pratt, J. (2008). Digits affect actions: the SNARC effect and response selection. *Cortex, 44(4),* 400–405.

Darwin, C. (1872). *The Expression of Emotions in Animals and Man*. London: Murray.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex, 48(7),* 882–887.

Dennis, I., & Newstead, S. E. (1981). Is phonological recoding under strategic control? *Memory & Cognition, 9(5),* 472–477.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284–311). Cambridge: Cambridge University Press.

Devlin, J. T., & Aydelott, J. (2009). Speech perception: Motoric contributions versus the motor theory. *Current Biology, 19(5),* R198–R200.

Desmurget, M., Richard, N., Harquel, S., Baraduc, P., Szathmari, A., Mottolese, C., & Sirigu, A. (2014). Neural representations of ethologically relevant hand/mouth synergies in the human precentral gyrus. *Proceedings of the National Academy of Sciences, 111(15)*, 5718–5722.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91(1),* 176–180.

Eiter, B. M., & Inhoff, A. W. (2008). Visual word recognition is accompanied by covert articulation: Evidence for a speech-like phonological representation. *Psychological Research, 72(6),* 666–674.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology, 91(4),* 451–471.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15(2),* 399–402.

Fagan, M. K., & Iverson, J. M. (2007). The influence of mouthing on infant vocalization. *Infancy, 11(2),* 191–202.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17(8),* 1703–1714.

Flöel, A., Ellger, T., Breitenstein, C., & Knecht, S. (2003). Language perception activates the hand motor cortex: implications for motor theories of speech perception. *European Journal of Neuroscience, 18(3),* 704–708.

Fogel, A., & Hannan, T. E. (1985). Manual actions of nine-to fifteen-week-old human infants during face-to-face interaction with their mothers. *Child Development, 56*, 1271–1279.

Futagi, Y., Yanagihara, K., Mogami, Y., Ikeda, T., & Suzuki, Y. (2013). The Babkin reflex in infants: clinical significance and neural mechanism. *Pediatric Neurology, 49(3),* 149–155.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119(2),* 593–609.

Gentilucci, M. (2003). Grasp observation influences speech production. *European Journal of Neuroscience, 17(1),* 179–184.

Gentilucci, M., Benuzzi, F., Bertolani, L., Daprati, E., & Gangitano, M. (2000). Language and motor control. *Experimental Brain Research, 133(4),* 468–490.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology, 86(4),* 1685–1699.

Gentilucci, M., & Campione, G. C. (2011). Do postures of distal effectors affect the control of actions of other distal effectors? Evidence for a system of interactions between hand and mouth. *PLOS One, 6(5),* e19793.

Gentilucci, M., Campione, G. C., Dalla Volta, R., & Bernardis, P. (2009). The observation of manual grasp actions affects the control of speech: A combined behavioral and Transcranial Magnetic Stimulation study. *Neuropsychologia, 47(14),* 3190–3202.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews, 30(7),* 949–960.

Gentilucci, M., Santunione, P., Roy, A. C., & Stefanini, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *European Journal of Neuroscience, 19(1),* 190–202.

Glover, S. (2004). Separate visual representations in the planning and control of action. *Behavioral and Brain Sciences, 27(1),* 3–24.

Glover, S., & Dixon, P. (2002). Semantics affect the planning but not control of grasping. *Experimental Brain Research, 146(3),* 383–387.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14(4),* 531–551.

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? *Studies in the Evolution of Language, 3,* 158–181.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology,14,* 5–24.

Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes, 25(6),* 749–776.

Higginbotham, D. R., Isaak, M. I., & Domingue, J. N. (2008). The exaptation of manual dexterity for articulate speech: an electromyogram investigation. *Experimental Brain Research, 186(4),* 603–609.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders, 36(2),* 149–171.

Hommel, B. (1996). SR compatibility effects without response uncertainty. *The Quarterly Journal of Experimental Psychology Section A, 49(3),* 546–571.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37(2),* 229–261.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies, 6(11-12),* 19–40.

Komeilipoor, N., Tiainen, M., Tiippana, K., Vainio, M., & Vainio, L. (2016). Excitability of hand motor areas during articulation of syllables. *Neuroscience Letters, 620,* 154–158.

Ladefoged, P., (2001). *Vowels and Consonants: An Introduction to the Sounds of Language*, Maldern, MA: Blackwell.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review, 74(6),* 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21(1),* 1–36.

Loetscher, T., Bockisch, C. J., Nicholls, M. E., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology, 20(6),* R264–R265.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide.* New York: Cambridge University Press.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21(4),* 499–511.

McGuigan, F. J. (1970). Covert oral behavior during the silent performance of language tasks. *Psychological Bulletin, 74(5),* 309–326.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., & Töpper, R. (2003). Motor cortex hand area and speech: Implications for the development of language. *Neuropsychologia, 41(4),* 401–406.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17(19),* 1692–1696.

Mochida, T., Kimura, T., Hiroya, S., Kitagawa, N., Gomi, H., & Kondo, T. (2013). Speech misperception: speaking and seeing interfere differently with hearing. *PLOS One, 8(7),* e68619.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20(8),* 750–756.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research, 13(3),* 417–425.

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience, 29(31),* 9819–9825.

Napier, J. R. (1956). The prehensile movements of the human hand. The Journal of bone and joint surgery. *British Volume, 38(4),* 902–913.

Nelson, E. L., Campbell, J. M., & Michel, G. F. (2014). Early handedness in infancy predicts language ability in toddlers. *Developmental Psychology, 50(3),* 809–814.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385(6615),* 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology, 118(12),* 2544–2590.

Paget, R. (1930) *Human Speech*. New York: Harcourt Brace.

Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *Neuroimage, 14(1),* 182–193.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11(5),* 351–360.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103(20),* 7865–7870.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21(5),* 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169–192.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research, 71(3),* 491–507.

Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology, 25(6),* 871–884.

Rosenbaum, D. A. (1980). Human movement initiation: specification of arm, direction, and extent. *Journal of Experimental Psychology: General, 109(4),* 444–474.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323(5916),* 951–953.

Salmelin, R., & Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Human Brain Mapping, 16(2),* 81–91.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127(1),* 141–145.

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23(2-3),* 429–435.

Sato, M., Troille, E., Ménard, L., Cathiard, M. A., & Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research, 227(2),* 275–288.

Smalle, E. H., Rogers, J., & Möttönen, R. (2014). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cerebral Cortex, 25(10),* 3690–3698.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73(4),* 971–995.

Sweet, H. (1888). *A History of English Sounds From the Earliest Period: With Full Word-Lists*. Oxford: Clarendon Press.

Symes, E., Tucker, M., Ellis, R., Vainio, L., & Ottoboni, G. (2008). Grasp preparation improves change detection for congruent objects. *Journal of Experimental Psychology: Human Perception and Performance, 34(4),* 854–871.

Tiainen, M., Felisberti, F. M., Tiippana, K., Vainio, M., Šimko, J., Lukavský, J., & Vainio, L. (2016). Congruency effect between articulation and grasping in native English speakers. In *Proceedings Interspeech, 2016,* (pp. 1108–1112). International Speech Communication Association.

Tiainen, M., Lukavský, J., Tiippana, K., Vainio, M., Šimko, J., Felisberti, F., & Vainio, L. (2017). Connections of grasping and horizontal hand movements with articulation in Czech speakers. *Frontiers in Psychology, 8,* 516.

Tiainen, M., Tiippana, K., Paavilainen, P., Vainio, M., & Vainio, L. (2017). Mismatch negativity (MMN) to speech sounds is modulated systematically by manual grip execution. *Neuroscience Letters, 651,* 237–241.

Tiainen, M., Tiippana, K., Vainio, M., Komeilipoor, N., & Vainio, L. (2017). Interaction in planning vocalizations and grasping. *The Quarterly Journal of Experimental Psychology, 70(8),* 1590–1602.

Tiainen, M., Tiippana, K., Vainio, M., Peromaa, T., Komeilipoor, N., & Vainio, L. (2016). Selective influences of precision and power grips on speech categorization. *PLOS One. 11(3),* e0151688.

Tokimura, H., Asakura, T., Tokimura, Y., Oliviero, A., & Rothwell, J. C. (1996). Speech-induced changes in corticospinal excitability. *Annals of Neurology, 40(4),* 628–634.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8(6),* 769–800.

Vainio, L. (2019). Connection between movements of mouth and hand: Perspectives on development and evolution of speech. *Neuroscience & Biobehavioral Reviews, 100,* 211–223.

Vainio, L., Schulman, M., Tiippana, K., & Vainio, M. (2013). Effect of syllable articulation on precision and power grip performance. *PLOS One, 8(1),* e53061.

Vainio, L., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2015). Interaction in planning movement direction for articulatory gestures and manual actions. *Experimental Brain Research, 233(10),* 2951–2959.

Vainio, L., Tiainen, M., Tiippana, K., & Vainio, M. (2014). Shared processing of planning articulatory gestures and grasping. *Experimental Brain Research, 232(7),* 2359–2368.

Van Hooff, J. A. R. A. M. (1967): The facial displays of catarrhine monkeys and apes. In D. Morris (Ed.), *Primate Ethology* (pp. 7–68). London: Weidenfeld & Nicolson.

Voelkl, B., & Huber, L. (2000). True imitation in marmosets. *Animal Behaviour, 60(2),* 195–202.

Wacewicz, S., Żywiczyński, P., & Orzechowski, S. (2016). Visible movements of the orofacial area. *Gesture, 15(2),* 250–282.

Wallace, A. R. (1881). Review of: Anthropology, by Edward B. Tylor. *Nature, 24,* 242–245.

Waters, G. S., & Fouts, R. S. (2002). Sympathetic mouth movements accompanying fine motor movements in chimpanzees (Pan troglodytes) with implications toward the evolution of language. *Neurological Research, 24(2),* 174–180.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41(8),* 989–994.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101(3),* 278–289.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology, 21(3–4),* 147–163.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception, 44(3),* 289–300.

Woll, B. (2014). Moving from hand to mouth: echo phonology and the origins of language. *Frontiers in Psychology, 5,* 662.